Modular Forms and Lattice Point Counting Problems

Carlos Pastor Alcoceba

Supervised by: Fernando Chamizo Lorente

A thesis submitted for the degree of
Doctor of Mathematics

Instituto de Ciencias Matemáticas
Universidad Autónoma de Madrid
Departamento de Matemáticas

Abstract

The results contained in this dissertation correspond to three problems lying in the interface between harmonic analysis and analytic number theory, which are described below. The memoir can be divided in two parts: the first three chapters discuss the problem of determining the regularity of fractional integral of classical modular forms, while the latter three deal with lattice point counting problems.

Given a nonzero modular form f, we can expand it in the Fourier series $f(z) = \sum_{n+\kappa\geq 0} e^{2\pi i(n+\kappa)z/m}$. We can associated to this series for every $\alpha \in \mathbb{R}$ the formal series $f_\alpha(x) = \sum_{n+\kappa>0} e^{2\pi i(n+\kappa)x/m}$ which converges for α big enough to a continuous function on the real line. The problem of determining the regularity of the resulting functions when this procedure is applied to Jacobi's theta function $\theta(z) = \sum_{n\in\mathbb{Z}} e^{\pi i n^2 z}$ has attracted a great deal of interest since Weierstrass presented one of these functions as an example given by Riemann of a nowhere differentiable function [6, 7, 8, 9, 10, 11]. The investigation of the regularity of functions obtained from arbitrary modular forms has only been started recently [1, 5, 16]. In particular, an interesting problem consisted in determining the pointwise Hölder exponent $\beta(x)$ defined by

$$\beta(x_0) = \sup\big\{s \ : \ \text{for some polynomial } P \text{ one has } f(x)-P(x) = O\big(|x-x_0|^s\big)\big\}.$$

Another interesting problem consisted in determining the spectrum of singularities, *i.e.* determining the Hausdorff dimension of those sets where the function attains a particular pointwise Hölder exponent. In both directions there were many open questions, which have been essentially solved in chapter 3 of this memoir. More concretely, we provide explicit formulas to determine the exponent $\beta(x)$ associated to f_α, which in some cases depends on some Diophantine approximation properties of the real x; and we provide the graph of the spectrum of singularities. We also determine an approximate functional equation, similar to the one obtained by Duistermaat [6] for the aforementioned "Riemann's example".

In the second half of the thesis we discuss lattice point counting problems. Given a convex d-dimensional set K, we are interested in estimating the error exponent

$$\alpha_K = \inf\big\{\alpha \ : \ \mathcal{N}_K(R) - R^d|K| = O\big(R^\alpha\big)\big\}$$

where $|K|$ denotes the volume of K and $\mathcal{N}_K(R)$ the number of points with integer coordinates lying inside K after having applied a homothety fixing the origin of ratio R. In general the problem of determining α_K for a particular convex body K is difficult and it has only been solved in very few cases. For

example, the most paradigmatic case, "Gauss circle problem", where K is a circle in the plane, is still open. In the chapter 5 we prove that $\alpha_K \leq 1$ when K is the double revolution paraboloid given by $\left\{|y| \leq c - (x_1^2 + x_2^2)\right\}$, result which is also extended to a family of paraboloids having rational ellipses as bases. This generalizes a result obtained by Popov for the parabola [15] and improves some previous results given by Krätzel [12, 13]. This result is expected to be sharp. The approach used to obtain these results consists in applying the Poisson summation formula to translate the problem into the estimation of an exponential sum, which happens to be related in this case to Jacobi's theta function. We can then employ a toy version of the circle method to obtain the needed bounds. In the same chapter we also provide lower bounds for α_K under more restrictive hypothesis, proving that indeed $\alpha_K = 1$ in many cases.

Finally in chapter 6 we consider three-dimensional convex bodies which are of revolution and have smooth boundary with nonvanishing Gaussian curvature. For these objects Chamizo proved in [2] that one has $\alpha_K \leq 11/8$ provided that the third derivative of the generatrix does not vanish anywhere. In this chapter we prove that the latter hypothesis can be weakened and the result holds when all the zeros of the generatrix are of finite order. To obtain this result we basically use the van der Corput method, although at some point one has to consider Diophantine properties of the phase function which bear some similarity with the case of the paraboloid trated in the previous chapter.

The results of the chapters 3, 5 and 6 correspond to the articles [14], [3] and [4]. The other three chapters contain introductory material.

Resumen

Los resultados contenidos en esta tesis corresponden a tres problemas que quedan en la interfaz entre el análisis armónico y la teoría analítica de números, y que se describen abajo. La tesis en si se puede considerar dividida en dos partes: los primeros tres capítulos tratan sobre la regularidad de las integrales fraccionarias de formas modulares clásicas, y los tres últimos sobre problemas de conteo de puntos del retículo.

Dada una forma modular no nula f, esta admite una expansión en serie de Fourier $f(z) = \sum_{n+\kappa\geq 0} e^{2\pi i(n+\kappa)z/m}$. A esta se le puede asociar para $\alpha \in \mathbb{R}$ la serie formal $f_\alpha(x) = \sum_{n+\kappa>0} e^{2\pi i(n+\kappa)x/m}$, que converge a una función continua en toda la recta real para α suficientemente grande. Determinar la regularidad de las funciones que resultan de aplicar esta construcción a la función theta de Jacobi $\theta(z) = \sum_{n\in\mathbb{Z}} e^{\pi i n^2 z}$ es un problema que ha generado un gran volumen de bibliografía a raíz de que Weierstrass presentara una de estas funciones como un ejemplo dado por Riemann de una función no diferenciable en ningún punto [6, 7, 8, 9, 10, 11]. Las funciones obtenidas a partir de formas modulares arbitrarias han empezado a ser consideradas para su estudio recientemente [1, 5, 16]. En particular, era objeto del deseo determinar el llamado exponente Hölder puntual $\beta(x)$ definido por

$$\beta(x_0) = \sup\big\{s \ : \ \text{existe un polinomio } P \text{ tal que } f(x)-P(x) = O\big(|x-x_0|^s\big)\big\}.$$

También era interesante determinar el espectro de singularidades, compuesto por las dimensiones de Hausdorff de aquellos conjuntos donde la función alcanza un exponente Hölder puntual en particular. En ambas direcciones quedaban muchas cuestiones abiertas, que quedan esencialmente resueltas en el capítulo 3 esta tesis. Más concretamente, se dan fórmulas para determinar el exponente $\beta(x)$ asociado a f_α, que en ciertos casos tiene que ver con temas de aproximación diofántica del real x en cuestión; y se da el grafo del espectro de singularidades. Además se determina una ecuación funcional aproximada, al estilo de la hallada por Duistermaat [6] para el arriba mencionado "ejemplo de Riemann".

En la segunda mitad de la tesis se tratan problemas de conteo de puntos del retículo. Dado un cuerpo convexo d-dimensional K, estamos interesados en estimar el exponente de error

$$\alpha_K = \inf\big\{\alpha \ : \ \mathcal{N}_K(R) - R^d|K| = O\big(R^\alpha\big)\big\}$$

donde $|K|$ denota el volumen de K y $\mathcal{N}_K(R)$ el número de puntos de coordenadas enteras que caen dentro de K después de haber sido dilatado por una

homotecia fijando el origen de razón R. En general, una vez fijado K, determinar α_K es un problema difícil que sólo ha podido resolverse en algunos casos particulares. Por ejemplo, el caso más paradigmático, el "problema del círculo de Gauss", cuando K es un círculo en el plano, aún está abierto. En el capítulo 5 se prueba $\alpha_K \leq 1$ cuando K es el doble paraboloide de revolución determinado por $\left\{|y| \leq c - (x_1^2 + x_2^2)\right\}$, resultado que se extiende a una familia de paraboloides cuyas bases son elipses racionales. Esto extiende un resultado obtenido por Popov para para parábola [15] y mejora resultados previos de Krätzel [12, 13]. Este resultado se espera que sea óptimo. La estrategia seguida para obtener estos resultados consiste en emplear sumación de Poisson para sustituir el problema por el de acotar una suma exponencial, que en este caso está relacionada con la función theta de Jacobi. Para obtener las cotas necesarias se emplea una versión simplificada del método del círculo. En el mismo capítulo también se dan cotas inferiores para α_K bajo hipótesis ligeramente más fuertes, mostrando que efectivamente $\alpha_K = 1$ en muchos casos.

Finalmente en el capítulo 6 se tratan cuerpos convexos de revolución tridimensionales con frontera suave con curvatura de Gauss no nula. Para estos objetos Chamizo probó en [2] que se tiene $\alpha_K \leq 11/8$ pidiendo que la tercera derivada de la función generatriz no se anulara en ningún punto. En este capítulo se prueba que basta con pedir que dicha generatriz, si se anula, lo haga con ceros de orden finito. Para esto se utiliza esencialmente el método de van der Corput, aunque en cierto punto hay que involucrar propiedades diofánticas de la fase que recuerdan al problema del paraboloide tratado en el capítulo anterior.

Los resultados de los capítulos 3, 5 y 6 corresponden a los artículos [14], [3] y [4]. Los otros tres capítulos contienen material introductorio.

References

[1] F. Chamizo. *Automorphic Forms and Differentiability Properties.* Trans. Amer. Math. Soc., 356(5):1909–1935 (electronic), 2004.

[2] F. Chamizo. *Lattice points in bodies of revolution.* Acta Arith., 85(3):265–277, 1998.

[3] F. Chamizo, C. Pastor. *Lattice points in elliptic paraboloids.* arXiv:1611.04498v2, 2017 (to appear in Publicacions Matemàtiques).

[4] F. Chamizo, C. Pastor. *Lattice points in bodies of revolution II.* arXiv:1709.08593v2, 2017.

[5] F. Chamizo, I. Petrykiewicz, S. Ruiz-Cabello. *The Hölder exponent of some Fourier series.* J. Fourier Anal. Appl., 23(4):758–777, 2017.

[6] J. J. Duistermaat. *Selfsimilarity of "Riemann's Nondifferentiable Function".* Nieuw Arch. Wisk., 9(3):303–337, 1991.

[7] J. Gerver. *The differentiability of the Riemann function at certain rational multiples of π.* Amer. J. Math., 92:33–55, 1970.

[8] J. Gerver. *More on the differentiability of the Riemann function.* Am. J. Math., 93(1):33–41, 1971.

[9] G. H. Hardy. *Weierstrass's nondifferentiable function.* Trans. Amer. Math. Soc., 17(3):301–325, 1916.

[10] S. Jaffard. *The spectrum of singularities of Riemann's function.* Rev. Mat. Iberoamericana, 12(2):441–460, 1996.

[11] S. Jaffard. *Local behavior of Riemann's function.* In *Harmonic analysis and operator theory (Caracas, 1994)*, volume 189 of *Contemp. Math.*, pp. 287–307. Amer. Math. Soc, 1995.

[12] E. Krätzel. *Lattice points in elliptic paraboloids.* J. Reine Angew. Math., 416:25–48, 1991.

[13] E. Krätzel. *Weighted lattice points in three-dimensional convex bodies and the number of lattice points in parts of elliptic paraboloids.* J. Reine Angew. Math., 485:11–23, 1997.

[14] C. Pastor. *On the regularity of fractional integrals of modular forms.* arXiv:1603.06491, 2016 (to appear in Trans. of the Amer. Math. Soc.).

[15] V. N. Popov. *The number of lattice points under a parabola.* Mat. Zametki, 18(5):699–704, 1975.

[16] S. Ruiz-Cabello. *Generadores de primos, identidades aproximadas y funciones multifractales.* PhD dissertation, Universidad Autónoma de Madrid, 2014.

A mis padres, ante todo;
pues es mérito suyo.

Contents

Foreword

Dies diem docet

This dissertation, dear reader, is the reflection of the journey that a PhD represents. It can therefore be seen as a kind of journal, where the material, the difficulties found along the way and their corresponding workarounds are presented more or less in the chronological order they were encountered. In an attempt to make the journey as enjoyable as it was for me, the ideas are presented from the simplest to the most complex —as is often the way in which they naturally arise in the mathematician mind when stepping into *terra incognita*. Following this principle we will take a slight detour whenever possible to discuss the most paradigmatic case: simple, transparent, yet sharing the main difficulties with the general case, before engaging in meaningless technicalities.

Along the formal proofs I have also tried to pack all the intuition I have developed about the topic being considered, with the intention it could serve as a map to others starting their own journey. Hopefully this will become common practice in the near future, as mathematics is not only about theorems and rigor, but also about ideas and intuition.

Introduction:
Two tales connected to Jacobi's theta function

The original objective proposed for this dissertation was to solve several small but interesting problems, sharing the common feature that they lie in the intersection between analytic number theory and harmonic analysis. If we had however to choose a leitmotif *a posteriori* for the whole exposition it would definitely be Jacobi's theta function

$$\theta(z) = \sum_{n\in\mathbb{Z}} e^{\pi i n^2 z}. \tag{I.1}$$

This function, clearly holomorphic in the upper half-plane by virtue of the uniform convergence on compact sets, is intimately linked to the arithmetic properties of the sequence of squares $\{n^2\}$ of the integer numbers. But this was not the main reason why Jacobi studied it, as he was originally concerned with the theory of elliptic integrals. In fact, he actually defined a more general function Θ, depending on two complex variables, of which θ is only a particular case. For our purposes, however, θ as defined in (**I**.1) will suffice, and therefore we will keep this notation throughout this document. In the following section the interested reader will find some brief notes about the original work of Jacobi. Later on we will provide a historical introduction to the problems addressed in this dissertation.

I.1. Historical remarks

After seeing the derivation of the equation for the pendulum in high school I remember being intrigued by the fact that the small angle approximation $\sin x \approx x$ seems unavoidable if one desires to obtain a closed expression for the law governing its movement. Indeed, suppose we have a pendulum of length ℓ and denote by $\nu(t)$ the angle from the vertical to the string at time t. Newton's law $F = ma$ then translates to the differential equation

$$\ell\nu''(t) + g\sin\nu(t) = 0,$$

where g denotes the acceleration due to gravity. If we multiply the equation by $2\nu'$ and integrate from 0 to t, we obtain

$$\ell\big(\nu'(t)\big)^2 - 2g\cos\nu(t) = -2g\cos\nu_0. \tag{I.2}$$

We have named $\nu_0 = \nu(0)$ the initial angle, and we have also assumed the pendulum is not moving at time zero, *i.e.* $\nu'(0) = 0$. Equation (**I**.2) only determines ν' up to sign, but physical intuition tells us that its sign has to be negative if $\nu_0 > 0$, at least for the first half-period, and therefore we must have

$$\nu'(t) = -\sqrt{2g\ell^{-1}\big(\cos\nu(t) - \cos\nu_0\big)}.$$

Since the variables are separated, and it is reasonable to assume that ν is injective in each half-period, inverting the relationship between ν and t we may write

$$-t\sqrt{2g\ell^{-1}} = \int_{\nu_0}^{\nu} \frac{du}{\sqrt{\cos u - \cos \nu_0}}. \tag{I.3}$$

At this point, however, we are stuck. No matter what we try it seems impossible to solve the integral —and indeed it is.[1] But a rigorous proof of this fact is out of the scope of this exposition. Let us ignore this fact for now, and perform anyway the change of variables $\sin(u/2) = v$, reminiscent of the tangent of the half-angle change of variables we were once taught as magically solving any integral involving trigonometric functions. Writing $k = \sqrt{(1-\cos\nu_0)/2} = \sin(\nu_0/2)$ and $v = kw$, equation (**I**.3) is then equivalent to

$$t\sqrt{g\ell^{-1}} = \int_{k^{-1}\sin(\nu/2)}^{1} \frac{dw}{\sqrt{(1-w^2)(1-k^2w^2)}}. \tag{I.4}$$

It is convenient at this point to deviate briefly from the case of the pendulum and consider instead the general case of the indefinite integral

$$\int_c^x R\Big(t, \sqrt{P(t)}\Big)\, dt \tag{I.5}$$

where c is a constant, R is a rational function and P a polynomial. Note (**I**.4) provides a particular example of an integral of this kind where P is a polynomial of fourth degree. If P had degree lower than three then we would have no problem solving the integral. Indeed, if P is constant then the integrand reduces to a rational function, and we know we can always express the integral in a closed form by means of the logarithm $\int t^{-1}dt$ and the arctangent[2] $\int(1+t^2)^{-1}dt$ functions. When $\deg P = 1$ or 2 essentially no new functions appear: in the first case the change of variables $v^2 = P(t)$ reduces the integrand again to a rational function, while in the second case we may complete squares to assume either $P(t) = 1 - t^2$ or $P(t) = 1 + t^2$. We may then perform the change of variables $t = \sin u$ and $\tan(u/2) = v$ (or its hyperbolic analogue) to reduce the integrand to a rational function. Note the relationship between t and v is in both cases algebraic, ensuring the result is always a composition of logarithm, arctangent and algebraic functions.

When $\deg P \geq 3$ however this is no longer the case, and new transcendental functions are required to express (**I**.5) in a closed form. The cases $\deg P = 3$ and 4 are very alike and particularly interesting, as these integrals appear in a natural way in several classical problems. These include the computation of the arc-length of an ellipse as a function of the angle, the distance to the Sun of a planet as a function of time or the evolution of the pendulum, as we already know by (**I**.4). From the first of these problems, the integral (**I**.5) borrows the name of *elliptic integral* when P is any cubic or quartic polynomial, and the particular case

$$F(x;k) = \int_0^x \frac{dt}{\sqrt{(1-t^2)(1-k^2t^2)}}, \tag{I.6}$$

incomplete elliptic integral of the first kind. The family of functions $F(x;k)$ (depending on the parameter k, which receives the name of *modulus*) together with two

[1] This means there is no closed formula representing the integral in terms of the variable ν and involving only elementary functions: rational functions (or even algebraic functions), exponential, logarithmic and trigonometric functions.

[2] If we allow the use of complex numbers then the logarithm suffices, as $\arctan x = (2i)^{-1}\log(x-i)/(x+i)$.

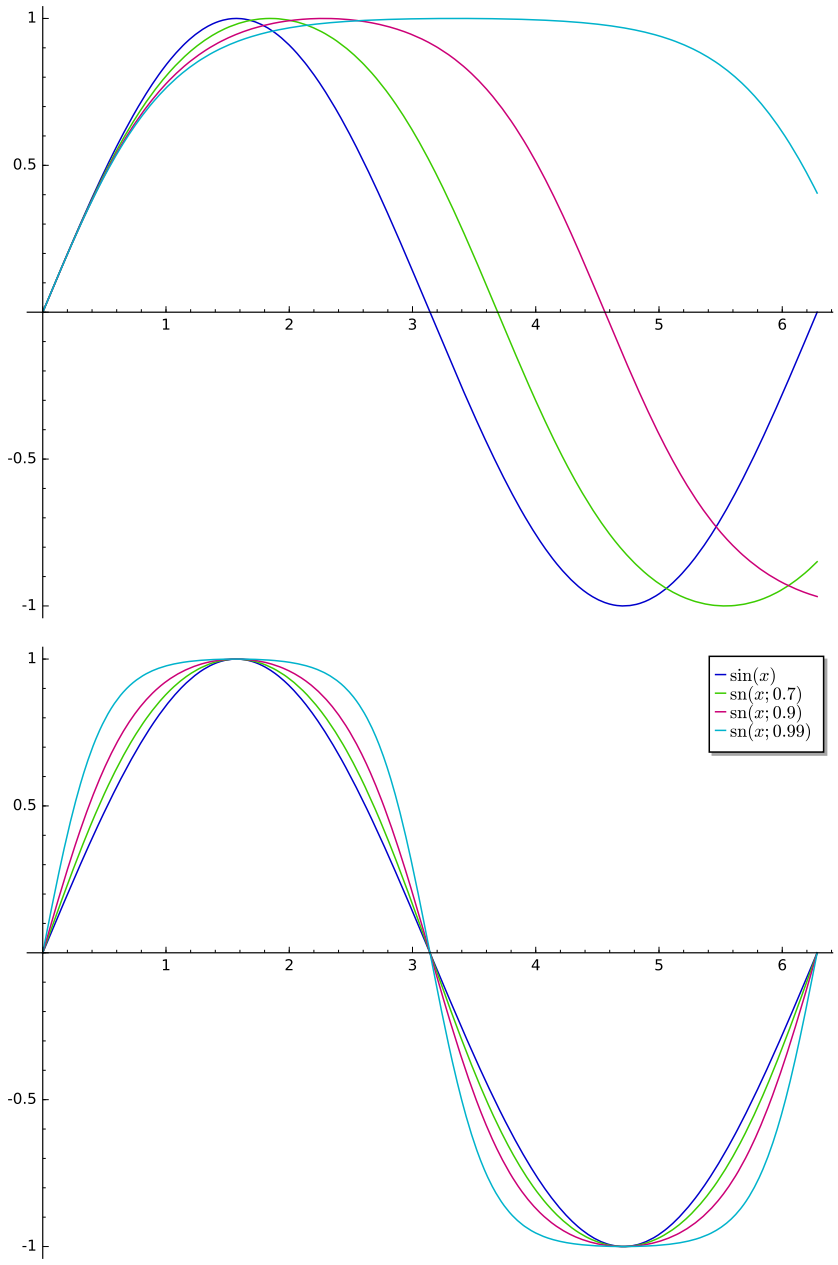


FIGURE I.1. The elliptic sine function for several values of the modulus k. In the bottom image the x variable has been reescaled for each value of k to make all periods match 2π.

families more, namely the elliptic integrals of the second and third kinds (which will not be introduced here), suffice to express any elliptic integral (I.5) in a closed form.

It turns out it is much easier to study the inverse function of $F(\,\cdot\,;k)$ than to study F itself. One of the reasons is that if one tries to employ properties of the integral (I.6) to extend its domain of definition, one ends up with a multivaluated function. This is analogous to what happens with the logarithm or the arcsine functions, which may also be defined as the integrals $\int t^{-1}dt$ or $\int(1-t^2)^{-1/2}dt$, but it is often easier to define the exponential or sine functions first, study their properties and then translate them to their inverses. Jacobi noticed this and, after

the work of Legendre and Abel, studied and named the inverse of F *elliptic sine*, abbreviated sn and by definition satisfying $F(\operatorname{sn}(x,k);k) = x$. With this notation we may rewrite (**I**.4) as

$$\sin(\nu/2) = -k \operatorname{sn}\left(\sqrt{g\ell^{-1}}(t-t_0), k\right) \tag{**I**.7}$$

for a certain constant t_0. Equation (**I**.7) essentially solves the problem we had at hand: providing a "closed" expression for the law describing how the pendulum evolves. Of course, introducing the elliptic sine in the syllabus of a high school course just to derive this formula would make little sense, but nevertheless (**I**.7) could still be mentioned to discuss the behavior of the pendulum, at least when the initial angle ν_0 is large. Figure **I**.1 shows the aspect of sn for several values of the modulus k. For example, in the figure it can be seen that the period of the pendulum is not truly constant, but depends on the initial angle ν_0, and in fact tends to infinity when $|\nu_0|$ approaches π (for $|\nu_0| > \pi/2$ the string has to be replaced by a rigid rod for the experimental setup to make sense).

Note we can deduce from (**I**.6) that for $k = 0$ the elliptic sine coincides with the usual sine function, recovering the classic law for the pendulum when the angle is small. In fact, even when $k \neq 0$ the elliptic sine shares many features with the usual sine. It also has a companion, the *elliptic cosine* cn, and together they satisfy many formulas which are analogous to the usual trigonometric relations.[3] In particular we have addition formulas similar to those determining the values of the sine and cosine functions for the sum of angles. Note that through the parametrization $x = \cos t$, $y = \sin t$ these formulas provide the usual group law for the unit circle. In the same way the elliptic trigonometric functions can be used to parametrize a curve and define a group law over it. These curves receive the name of *elliptic curves*, and are of an outstanding importance in contemporary number theory.[4]

The reader acquainted with the theory of elliptic curves over the complex numbers will remember that these are always conformally equivalent to a torus constructed by quotienting the complex plane by a discrete subgroup, generated by two linearly independent vectors (also known as a *lattice*). Meromorphic functions living on the elliptic curve can then be identified with meromorphic functions on $\mathbb{C}$ having two linearly independent complex periods. These functions receive the name of *elliptic functions*. It should not surprise the reader after the aforementioned connection between elliptic trigonometric functions and elliptic curves that the former are indeed elliptic functions: they admit meromorphic extensions to the whole complex plane with two periods, only one of which is real. In fact, the addition formulas can be used to carry on the addition law on the complex tori to the whole elliptic curve, extending it to include the image of the complex points, and in this way the elliptic functions provide not only a conformal map but also a group isomorphism.[5]

Elliptic functions can nevertheless be defined and studied with no reference to elliptic curves whatsoever, and have interest on their own. A simple application of Liouville's theorem shows that the only entire elliptic functions are the constants. This surprisingly simple fact provides a powerful tool to prove some deep relations

[3]It actually has two companions! The other one, the *elliptic delta* dn has no relevance for the usual trigonometry because $\operatorname{dn} \equiv 1$ when $k = 0$, but when $k \neq 0$ it irremediably appears intermingled with sn and cn as part of the elliptic trigonometric relations.

[4]The modern definition of an elliptic curve is the locus of real (or complex points) satisfying an equation of the form $y^2 = x^3 + ax + b$ for parameters a and b with $4a^3 + 27b^2 \neq 0$.

[5]Note this is also true for the usual trigonometric functions, which provide a group isomorphism from $(\mathbb{C}/\mathbb{Z}, +)$ to $(\mathbb{C}^*, \cdot)$ extending the one from $(\mathbb{R}/\mathbb{Z}, +)$ to $(\mathbb{S}^1, \cdot)$.

between *a priori* seemingly unrelated functions. Indeed: any two functions which are elliptic of the same periods and whose poles and zeros coincide —including multiplicity— must be a constant multiple of each other. This fact was exploited by Jacobi to construct alternative expressions for the elliptic trigonometric functions from which to study their properties and compute particular values. This is where the *Jacobi theta function* Θ comes into play. This function is defined by the following series:

$$\Theta(z;\tau) = \sum_{n\in\mathbb{Z}} q^{n^2} e^{2\pi i n z} \qquad \text{where} \quad q = e^{i\pi\tau}.$$

For a fixed τ in the upper half-plane it is entire in the z variable and satisfies

$$\Theta(z+1;\tau) = \Theta(z;\tau) \qquad \text{and} \qquad \Theta(z+\tau;\tau) = q^{-1}e^{-2\pi i z}\Theta(z;\tau),$$

i.e. it has a real period and "almost" a second complex one. These identities follow by rearranging the series, which is possible due to the absolute convergence. As a consequence, the quotient $\Theta(z+\tau/2;\tau)/\Theta(z+(\tau+1)/2;\tau)$ is an elliptic function of periods 1 and 2τ. One can now perform a dilation in the z variable and adjust τ to match the periods with those of the elliptic sine function. It can then be seen that all the poles and zeros align, and therefore, multiplied by an appropriate constant, this quotient provides another expression for sn. This and many other relations were provided by Jacobi in [**63**]. In fact, he proved that any elliptic function can be written as a linear combination of quotients of the function Θ and first derivatives of them. This general theory however has long been superseded by the conceptually simpler theory of Weierstrass, involving instead the function

$$\wp(z) = \frac{1}{z^2} + \sum_{n,m\in\mathbb{Z}} \left(\frac{1}{(z+n+2\tau m)^2} - \frac{1}{(n+2\tau m)^2} \right).$$

Weierstrass showed that any elliptic function can be written in an unique way in the form $G\big(\wp(z)\big) + \wp'(z)H\big(\wp(z)\big)$ where G and H are rational functions. The Weierstrass's $\wp$-function can also be used to parametrize and define the group law on elliptic curves, and this is often the approach chosen in modern treatises, such as Koblitz's [**69**].

Jacobi also found in the rigidity of elliptic functions, and in particular in the machinery of theta functions, a useful tool to prove some surprising number-theoretic identities. In this way he obtained his famous four-square theorem:

THEOREM (JACOBI). *The number of ways of representing an integer n as a sum of four squares is exactly eight times the sum of its divisors if n is odd, and twenty four times the sum of its odd divisors if n is even.*

To illustrate the relation to theta functions, note that the coefficients of $\big(\Theta(0;\tau)\big)^4$, considered as a power series in the variable q, are precisely the number of ways of writing each integer as a sum of four squares. We can then build another power series in the variable q, whose coefficients are precisely the sums of divisors prescribed by the statement of the theorem, and try to show that both power series must coincide. The problem is that neither of these functions depend on the variable z, in which they "should" be elliptic, and filling this gap requires ingenuity. Nowadays we know that it is easier to focus instead on the law by which both functions transform in the variable τ, and use this to show they must be equal. The function $\Theta(0;\tau)$, which coincides with $\theta(\tau)$ as defined in (**I**.1), turns out to be a *modular form* in the variable τ. Although this notion will rigorously defined in chapter 2, let us say

for now that this means that θ satisfies the transformation laws $\theta(\tau+1) = \theta(\tau)$ and $\theta(-1/(4\tau)) = \sqrt{-2i\tau}\theta(\tau)$. The important fact is that the vector space of all modular forms which transform in the same way is finite-dimensional, and we have effective bounds on its dimension. Therefore the proof of Jacobi's four-square theorem reduces to proving that the two power series in q are modular forms of the same kind (one of them being θ^4), and then computing a finite number of coefficients to check they are equal.

In this exposition neither elliptic integrals nor elliptic functions will play any role, but modular forms definitely will. One final historical remark about their origin. When writing the elliptic sine as a quotient of theta functions, the variable τ depends on the modulus k. After inverting this relation, the function $k(\tau)$ turns out to be a modular form, and this is the reason they bear the adjective modular.

I.2. Riemann's example

In 1872 Weierstrass presented a lecture in the Berlin Academy of Sciences on the topic of function continuity and differentiability. The lecture started as follows:[6]

> Until recently, it has been generally accepted that a well-defined and continuous function of a real variable can only have a first derivative whose value is indeterminate or becomes infinitely large at isolated points. Even in the works of Gauss, Cauchy, Dirichlet there is to my knowledge no statement doubting this, even though these mathematicians were accustomed to being the strongest critics in their science. Only Riemann, as I heard from some of his auditors, pronounced with certainty (in the year 1861, or perhaps even earlier) that this assumption is incorrect and is for example disproven by the function represented by the infinite series
>
> $$\sum_{n=1}^{\infty} \frac{\sin(n^2 x)}{n^2}.$$
>
> Unfortunately, the proof by Riemann has not been published, and does not appear either in his publications or through oral communications. This is all the more regrettable, as I do not even know for sure how Riemann addressed this himself to his audience. Those mathematicians who, after Riemann's statement had become known in wider circles, considered the matter, seemed to believe (at least in their majority), that it is enough to prove the existence of functions that are not differentiable in any small interval. The existence of functions of this type can be easily proven, and I believe therefore that Riemann only had in mind functions with no derivative at any value of the argument. The proof that the given trigonometric series is a function of this kind seems quite difficult to me; however, one can easily construct continuous function of a real variable, for which one can prove with the easiest means, that no value of x gives a well-defined derivative.

[6]Many thanks to Corentin Perret-Gentil for his help in this translation.

In the last sentence Weierstrass is obviously talking about his famous family of nowhere differentiable functions

$$\sum_{n\geq 0} a^n \cos(b^n \pi x), \tag{I.8}$$

for any choice of a, b satisfying $0 < a < 1$, b a positive odd integer and $ab > 1+3\pi/2$. The rest of the talk was focused on the properties of these functions and can be consulted in german in [**95**].

The claims made by Weierstrass on the function

$$\varphi(x) = \sum_{n\geq 1} \frac{\sin(n^2\pi x)}{n^2} \tag{I.9}$$

and its relation to Riemann are both surprising and unsettling. More so considering that no proof regarding its differentiability was published until half a century later, when Hardy in 1916 [**42**] developed a new method to study the differentiability of Weierstrass' function (**I**.8). The main idea can be sketched as follows: this function coincides with the real part of the complex function $\sum a^n e^{\pi i b^n z}$, holomorphic in the upper half-plane, and the growth of the derivative of the latter as the variable z approaches the real line is closely related to the differentiability of the former. The right tool to formalize this relation is a pair of abelian and tauberian theorems, as the decay introduced by the imaginary part of z regularizes the series in an analogous way as how Abel summation works. Hardy not only employs this machinery to give a new proof of the nowhere differentiability of (**I**.8), but also notices that the same method applies to other functions, most notably φ. In this case φ coincides with the imaginary part of $\sum n^{-2} e^{\pi i n^2 z}$, function which is essentially a primitive of Jacobi's theta function θ defined in (**I**.1). Hardy was therefore able to refer to a previous joint work with Littlewood [**45**] where they had studied the growth of θ near the real line, among other related questions. In this way he succeeds in giving the first (known) proof that the derivative of φ cannot exist in a dense set. In fact, the only points where he was not able to determine the nondifferentiability of φ were the rational numbers of the form odd/odd or even/$(4n+3)$.

Could this, or a similar proof, have been known to Riemann? Hardy probably was doubtful because, even though he does attribute the result to Riemann in [**42**], he explicitly quotes Du Bois-Reymond as his source, who in [**9**] asserted:

> For some years now, there has been much talk in the german mathematical circles of the existence of functions without derivatives, especially since Riemann's disciples have declared that their teacher claimed the non-differentiability of the series with term $\sin(p^2x)/p^2$. In any small interval there should be values of x for which this series admits no derivative. To the best of my knowledge none of the Riemann pupils procured proof of this, but according to a statement by Weierstrass, Riemann's assertion is correct.

This was written in 1874, two years after the lecture in the Berlin Academy of Sciences. The claim can therefore be suspiciously traced back to Weierstrass. Not only that, but a letter reproduced in [**13**] also shows that it was Weierstrass himself who pressed Du Bois-Reymond into including this remark in his paper. The letter, written by the former to the latter, includes the following fragment:

> First of all I would consider it expedient to mention explicitly that Riemann already in the year 1861 has pointed out to some of his attenders that the function given by the series $\sum_{n=1}^{\infty} \sin(n^2x)/n^2$ is a function of the type that does not possess a derivative, that he however has not revealed his proof to anyone, but has only mentioned occasionally that it could be extracted from elliptic function theory.

The matter is more carefully studied by Butzer and Stark in [**13**]. In particular they found some correspondence from 1865 between Christoffel and Prym regarding the question of the differentiability of the closely related series $\sum \cos(n^2x)/n^2$. Only the letters from Christoffel are preserved, and although the replies are lost it seems Prym attempted a proof of their nondifferentiability which did not convince Christoffel, showing that at the time no proof had been communicated to them by Riemann. In the same letters it is also mentioned that Christoffel had discussed the problem with Weierstrass, possibly originating the confussion. If Prym did or not discuss the matter with Riemann we do not know; neither if, supposing he did, Riemann did provide a formal proof or just some intuition about the topic. Paraphrasing the authors of [**13**], although none of the direct students of Riemann have any detectable connection with it, who else other than Riemann had the imagination to create such an intriguing example!

In any case, the function φ has become known in the literature as "Riemann's example of a nondifferentiable function" (or "Riemann's example" for short), and indeed Hardy already refers to it in this terms in [**42**]. Another half century would have to pass for someone to finally settle the question of its differentiability at the remaining points, namely those rationals of the form odd/odd or even/$(4n+3)$. This was done by Gerver who, to everybody's surprise, in 1969 proved that φ has derivative $-\pi/2$ at every rational of the form odd/odd [**35**]. Some months later he completed the picture, showing that φ is not differentiable at the rationals of the form even/$(4n+3)$ [**36**]. The fact that φ is differentiable at some points could be suspected from the aspect of its graph, shown in figure **I**.2, but of course plots this detailed were not available at the time.

These results by Gerver, the link to Jacobi's theta function, and probably also the mystery surrounding its relation to Riemann, sparkled in the last fifty years a remarkable amount of literature, regarding different aspects of the regularity of φ and that of closely related functions. For example, Hardy had already considered in his original paper [**42**] the functions $\sum_{n\geq 1} \sin(n^2\pi x)/n^{2\alpha}$ for various values of $\alpha > 1/2$. After replacing the sine by a complex exponential, these functions essentially correspond to "primitives" of the Jacobi theta function of order α. To give a concrete meaning to this when α is not an integer one can resort to the *Riemann-Liouville integral*:[7]

$$I^\alpha f(y) = -\frac{1}{\Gamma(\alpha)} \int_y^\infty f(t)(y-t)^{\alpha-1}\, dt. \tag{**I**.10}$$

This functional satisfies many of the properties one should expect from a "fractional" integral when evaluated at functions which are good enough, including the identities $I^\alpha I^\beta f = I^{\alpha+\beta} f$, $(I^1 f)' = f$ and $(I^\alpha f)' = I^{\alpha-1} f$ for $\alpha > 1$. To apply it to Jacobi's theta function, however, we run into the problem that θ is not well-defined on

[7] The Riemann-Liouville integral is usually defined as $(\Gamma(\alpha))^{-1} \int_c^y f(t)(y-t)^{\alpha-1}\, dt$ for a base-point c. We have chosen $c = +\infty$ for convenience.

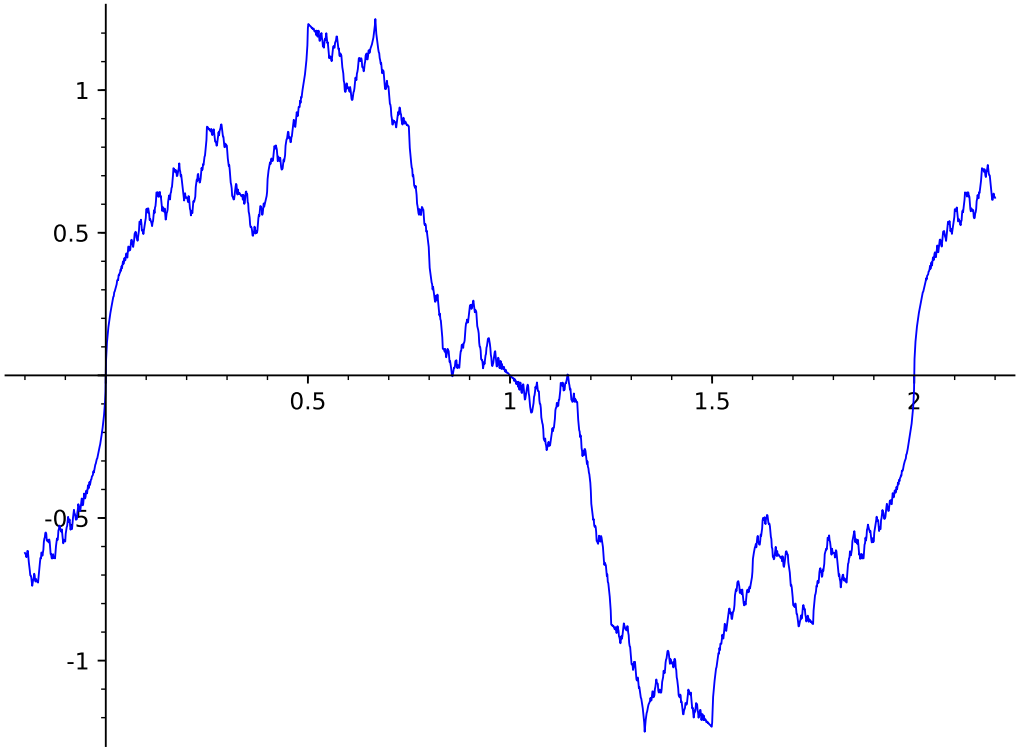


FIGURE I.2. The aspect of "Riemann's example" φ.

the real line. The solution is to apply it over a translated imaginary axis, after carefully removing the value $\lim_{t\to\infty}\theta(it) = 1$ to make it decay at infinity. The reader can check, assuming we may interchange integration and summation, that for $g_x(t) = \theta(x+it)-1$ we obtain $I^\alpha g_x(y) = C\theta_\alpha(x+iy)$, where $\theta_\alpha(z) = \sum_{n\geq 1} e^{\pi i n^2 z}/n^{2\alpha}$ and C is some constant depending on α. Hardy noticed this process can be inverted, essentially by applying $I^{-\alpha}$. To avoid problems with convergence, however, he had to replace the kernel of integration $(x-t)^{-\alpha-1}$ by a complex one. To illustrate this, consider the functional

$$J^\alpha f(z) = \int_{-\infty}^{+\infty} f(t)(t-z)^{-\alpha-1}\,dt \qquad \text{for} \quad \Im z > 0. \tag{I.11}$$

Note that by Cauchy's theorem the value of the following integral does not depend on y as long as $y > 0$:

$$\int_{-\infty-iy}^{+\infty-iy} e^{it}t^{-\alpha-1}\,dt.$$

Using this property the reader can also check, assuming again we may interchange integration and summation, that $J^\alpha\theta_\alpha(z) = C'(\theta(z)-1)$ for some constant C' depending on α.

We have plotted in figure I.3 the argument of the kernel $(t-z)^{-\alpha-1}$ for $\alpha = 1$ when $\Re z = 0$ for different values of $\Im z$ approaching 0. Note the graph remains almost constant except for $t \approx \Re z$, where the "sign" of the kernel undergoes a rapid variation, which is faster the smaller $\Im z$ is. Now, if f is very smooth around the point $x = \Re z$, the integral (I.11) will have extra cancellation in a neighbourhood of x, and as a result the divergence of $J^\alpha f(x+iy)$ when $y \to 0^+$ will be slower than if $|f|$ was integrated against $|t-z|^{-\alpha-1}$. If, on the contrary, f oscillates wildly

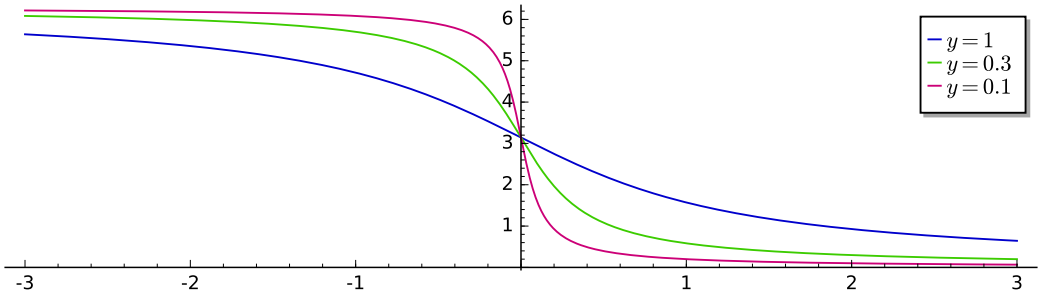


FIGURE **I**.3. A continuous determination of the argument of the kernel $(t-z)^{-\alpha-1}$ for $\alpha = 1$ and different values of $z = iy$. Other values of $\Re z$ translate the graph horizontally, while other values of α reescale it vertically.

around x, then for many small values of y it will resonate with the kernel, and the size of $J^\alpha f(x+iy)$ for small $y \approx 0$ will often resemble that obtained if $|f|$ was integrated against $|t-z|^{-\alpha-1}$. These heuristics are analogous to the fact that the smoother a function is the faster its Fourier transform decays. The advantage of this kernel over the complex exponential is that the oscillation is very localized, capturing information about *where* the function has low or high regularity. This is exactly what underlies the abelian an tauberian theorems exploited by Hardy in [**42**] to study "Riemann's example" φ and its generalizations θ_α. The same argument was later expanded by Holschneider, Tchamitchian, and Jaffard [**52, 64, 65**] allowing them to refine our knowledge on the regularity of these functions. At the time these three articles were published the aforementioned properties attributed to $(t-z)^{-\alpha-1}$ had already been studied for wide class of kernels, within the formalism of *wavelet transforms*. The wavelet transform of the function f with respect to the wavelet ψ is defined as:[8]

$$Wf(a,b) = \frac{1}{a}\int_{\mathbb{R}} f(t)\,\bar{\psi}\left(\frac{t-b}{a}\right)dt \qquad \text{for} \quad a>0 \quad \text{and} \quad b\in\mathbb{R}. \tag{I.12}$$

Note that if we take $\psi(t) = (t+i)^{-\alpha-1}$ then $Wf(y,x) = y^\alpha J^\alpha f(x+iy)$. In general the wavelet ψ must be a function which oscillates but at the same time has enough decay for the integral to converge. There is not a unique definition, and each author usually defines it as it is convenient for their purposes. For example, the axioms chosen by Holschneider and Tchamitchian [**52**], or by Jaffard [**64**], allowed them to prove different quantitative relations between the *Hölder continuity* of f at a point b and the decay of the transform $Wf(a,b)$ when $a \to 0^+$, generalizing the abelian and tauberian theorems originally provided by Hardy. Here Hölder continuity has to be understood in the following generalized sense: we say that a function is β-Hölder continous at x_0 for some $\beta > 0$ if there exists a polynomial p for which

$$f(x) = p(x-x_0) + O(|x-x_0|^\beta).$$

The supremum of all $\beta > 0$ for which f is β-Hölder continous at a point is then called the *Hölder exponent* of f at that point. Using this machinery these authors proved in [**52, 64, 65**] the following refinement of the theorems of Hardy and Gerver: φ

[8] In the words of Holschneider and Tchamitchian [**52**], "This transform is a sort of mathematical microscope, where $1/a$ is the enlargement and b is its position over the function to be analyzed. The specific optic is determined by the wavelet itself".

has Hölder exponent 3/2 at those rational numbers of the form odd/odd, while it has Hölder exponent 1/2 at the remaining rationals. They also tackled the question of the regularity at the irrational numbers, which is more subtle. At these points Hardy had already proved in [**42**] that the Hölder exponent of φ does not exceed 3/4. Jaffard in [**65**] was the first to compute this quantity precisely, showing that the following theorem holds:

THEOREM (JAFFARD). *Let x be an irrational number, and τ_x supremum of all the values of τ for which there exist infinitely many rationals p/q not of the form odd/odd satisfying $|x - p/q| \leq q^{-\tau}$. The Hölder exponent of φ at x then coincides with the quantity $1/2 + 1/(2\tau_x)$.*

The quantity τ_x can be regarded as a refinement of the usual notion of τ-approximability: an irrational number x is said to be *τ-approximable* if there are infinitely many rationals p/q for which $|x - p/q| \leq q^{-\tau}$. It is remarkable that the regularity of φ is so closely related to questions of Diophantine approximation! A classic theorem of Jarník and Besicovitch states that the Hausdorff dimension of the set of τ-approximable real numbers is precisely $2/\tau$ [**66**] (*cf.* [**7**]). Jaffard was able to extend this result to the set of τ-approximable numbers by rationals odd/odd, proving that the Hausdorff dimension of the set of points where φ has Hölder exponent β is $4\beta - 2$ for $\beta \in [1/2, 3/4]$. Functions for which the Hausdorff dimension of the sets $\{\text{Hölder exp.} = \beta\}$ may attain an infinite number of different values are refered to in the literature as *multifractal*, and they naturally arise in the study of turbulence [**33**]. In fact "Riemann's example" itself seems to be related to a special case of the evolution of a vortex filament equation [**54**].

So far we have neglected a key ingredient in all the aforementioned proofs: estimating the growth of Jacobi's theta function near the real line. The first authors to provide results in this direction were Hardy and Littlewood in [**45**], although the aim of this article was actually to study problems of Diophantine approximation. In a related previous article [**44**] they had studied whether given a polynomial p, an irrational number θ and any $\alpha \in [0, 1)$ one can find a sequence a_n of integers such that the fractional parts of $p(a_n)\theta$ converge to α. The answer is affirmative, and intimately related to the behavior of the family of exponential sums $\sum_{n \leq N} e^{2\pi i k p(a_n)\theta}$ indexed by k as $N \to \infty$. Their investigation was soon superseded by the beautiful criterion given by Weyl (see [**67**]), which we include here for delight of the reader:

THEOREM (WEYL'S CRITERION). *A sequence u_n of real numbers is equidistributed modulo 1 if, and only if, for all $k \in \mathbb{Z}^+$, $\frac{1}{N}\sum_{n=0}^{N} e^{2\pi i k u_n} \to 0$ as $N \to \infty$.*

Despite the generality of this result, Hardy and Littlewood had studied the case $p(x) = x^2$ in depth, and in particular the size of the quadratic exponential sums $\sum_{|n| \leq N} e^{\pi i n^2 x}$. When x is a rational p/q and $N = q$ these are the usual Gauss sums whose size was precisely determined by Gauss. When x is irrational, however, the question is not that simple. Hardy and Littlewood noticed that the size of the sum can be related to the growth of $|\theta(x + iy)|$ as $y \to 0^+$. This is, as in the previously presented case, because the truncated sum can be seen as a regularized version of $\theta(z)$, this time by sharply truncating the series instead of introducing the slowly decaying factor $1/n^{2\alpha}$. More importantly, they had the insight that a lot of information about the size of $|\theta(z)|$ can be obtained by ingeniously intermingling the functional equations $\theta(z+2) = \theta(z)$ and $\theta(-1/z) = \sqrt{-iz}\theta(z)$ in a way dictated by the continuous fraction expansion of the number $x = \Re z$. Although this can

be carried out as presented (using the functional equations to estimate $|\theta(z)|$ near the real line and then translating the result to bound the size of $\sum_{|n|\leq N} e^{\pi i n^2 x}$, *cf.* chapter 5) they found it easier to prove instead approximate analogues of the functional equations for the sums $\sum_{|n|\leq N} e^{\pi i n^2 x}$ with controlled error terms, and then infer directly from this their size.

Following a similar idea, Duistermaat succeeded in deriving an approximate functional equation for the function θ_1 from the one satisfied by θ, and was able to use this to extract more information about the behavior of "Riemann's example" φ. In the beautifully well-written article [**24**] he shows that the graph of φ, appropriately shrinked around a rational point and slightly modified by a differentiable error term, coincides again with itself. To illustrate the matter consider for a moment the function $f(x) = x\sin(2\pi/x)$ and note it satisfies the functional equation $f(x/(1+x)) = f(x)/(1+x)$, where the transformation $x/(1+x)$ fixes 0 and slightly shrinks or expands space around it. This forces f to oscillate wildly: indeed, any function satisfying this equation is of the form $xg(1/x)$ for some 1-periodic function g. A very similar argument shows that φ behaves like $Cx^{1/2} + x^{3/2}g(1/x)$ around every rational number, where the constant C and the periodic function g have to be chosen depending on the rational number. Duistermaat then goes on to show that the constant C is zero if an only if the rational is of the form odd/odd, and determines the possible functions g that may appear in this expansion. In fact, not only he provided new insight about the shape of the graph of φ around rational numbers, but he was also able to exploit the approximate functional equation to show, before Jaffard's theorem, that the Hölder exponent at irrational numbers is bounded above by $1/2 + 1/(2\tau_x)$.

In both approaches described above —the wavelet transform and the approximate functional equation one— the essential fact is that the function θ is a modular form. We will see in chapter 2 that every classical modular form satisfies a similar functional equation, and also admits a Fourier expansion essentially of the form $\sum_{n\geq 0} a_n e^{2\pi i n z}$. We can therefore construct the series $\sum_{n\geq 0} n^{-\alpha} a_n e^{2\pi i n x}$, which can be seen to converge to a continuous function for α big enough. This provides a source of very interesting Fourier series, which are only differentiable in certain subsets of the rational numbers when the parameter α is appropriately tuned, and satisfy approximate functional equations. The general study of these functions was initiated by F. Chamizo in [**14**], who determined for which ranges of α the Fourier series converge or diverge and characterized their differentiability under certain hypotheses.[9] One example extracted from the introduction of [**14**] is the following:

$$\sum_{n\equiv\pm 1 \pmod{12}} \frac{\sin(2\pi n^2 x)}{n^2} - \sum_{n\equiv\pm 5 \pmod{12}} \frac{\sin(2\pi n^2 x)}{n^2}. \tag{I.13}$$

This continuous function turns out to be only differentiable at the rational points, having null derivative at each of them. Other similar examples can be found in [**14**], together with some intriguing theorems relating the value of the derivative of these "fractional integrals" to arithmetic properties of the underlying modular forms. For example, the fact that the derivative of (**I**.13) vanishes at 0 is equivalent to the fact

[9] Roughly at the same time another article was published on the topic by Miller and Schmid [**77**]. They focus on the case $\alpha = 1$ for Maass forms —which are a non-holomorphic analogue of classical modular forms— but there is some overlapping with Chamizo's article [**14**]. Their approach is basically the same as the one employed by Duistermaat in [**24**].

that the L-function associated to the Dirichlet character χ modulo 12 determined by $\chi(\pm 1) = 1$ and $\chi(\pm 5) = -1$ also vanishes at 0.

The study of the regularity of these fractional integrals was later continued by Chamizo in a joint work with Petrykiewicz and Ruiz-Cabello [**19**], where they succeeded in computing the Hölder exponent only for very restricted ranges of α. Some further results with the same restrictions but involving some Diophantine analysis, which is essential to characterize the Hölder exponent at the irrational points, were also included in Ruiz-Cabello's PhD dissertation [**83**]. The weaknesses of their approach were the following: on the one hand they employed the same definition of wavelet as Jaffard, while a slightly modified definition proves more useful; and on the other hand they only provided a very rudimentary version of the approximate functional equation. Their approach is also restricted to a particular family of classical modular forms where the Diophantine analysis can be reduced to the notion of τ-approximability by rationals as employed by Jaffard in the theorem above, where the rationals are chosen from some congruence class. These deficiencies were addressed by the author in the article [**80**], with the inestimable help of F. Chamizo. The aforementioned techniques then are strong enough to prove analogues of the results of Jaffard and Duistermaat in the setting of arbitrary classical modular forms, and we devote chapter 3 of this dissertation to rigorously state and prove the theorems included in [**80**].

I.3. Gauss' circle problem

One of the main topics in Gauss' Disquisitiones Arithmeticae were *integral binary quadratic forms*. A quadratic form is a homogeneous polynomial of degree two, which is said to be binary if it depends on exactly two variables and integral if all the coefficients are integer numbers. We are therefore talking about objects of the form

$$Q(x,y) = ax^2 + bxy + cy^2 \qquad \text{where} \quad a,b,c \in \mathbb{Z}. \tag{I.14}$$

An equivalent, sometimes more convenient, way of representing the same object is as $Q(\vec{x}) = \vec{x}^t A \vec{x}$ where $A = \left(\begin{smallmatrix} a & b/2 \\ b/2 & c \end{smallmatrix}\right)$. In fact, for this reason, Gauss only considered those forms with even b so that the matrix A has integer coefficients, but nowadays it is common to let b be odd. We will offer in the next pages a glimpse of the general theory of integral binary quadratic forms.[10] The presented material is based on the exposition by Cohn [**22**]. For the sake of simplicity, the adjectives binary and integral will often be omitted.

When we evaluate an integral quadratic form in points with integer coordinates we obtain again integer values. Which integer values arise in this fashion for a given quadratic form is however a non-trivial problem. An even finer problem is to count in how many ways each integer can be obtained, if this quantity happens to be finite. To this end, we will say that the form Q represents n if the equation $Q(x,y) = n$ has an integer solution, and that it represents this integer k times if the number of integer solutions is exactly k. For example, the "simplest" form $Q(x,y) = x^2 + y^2$ represents 5 eight times, because $Q(\pm 1, \pm 2) = Q(\pm 2, \pm 1) = 5$ and there is no other way to obtain this integer. On the other hand it is easy to check that it never

[10] This may seem an exceedingly long disgression, however the author feels that the involved ideas, which lead in a natural way to the definition of the class number and to Gauss' circle problem, are too often omitted from number theory introductions.

represents 3. The general law underlying this phenomenon for this particular choice of Q was studied by Fermat and Euler, and can be summed up in the following two theorems:

THEOREM (GENUS). *The form $Q(x,y) = x^2 + y^2$ represents a prime p if and only if $p \equiv 1 \pmod 4$ or $p = 2$. The representation is unique except for obvious changes of sign and rearrangements of x and y.*

THEOREM (COMPOSITION). *The form $Q(x,y) = x^2 + y^2$ satisfies the composition law*

$$Q(x,y)\,Q(x',y') = Q(xx' - yy', x'y + xy') \tag{I.15}$$

and therefore if it represents integers n and m, it also represents their product nm. Moreover, every representation of an integer can be obtained by the composition law from either representations of prime numbers or from the trivial representations $Q(\pm p, 0) = Q(0, \pm p) = p^2$ of squares of prime numbers.

From these two facts we deduce that Q represents an integer if and only if all its prime divisors congruent to 3 modulo 4 appear in its factorization raised to an even power. The simplicity of these two theorems is due to the fact that the form $x^2 + y^2$ is in many ways special, but weaker variants hold true for all quadratic forms.

Some simplifications are convenient at this point. Note first that if all three coefficients of the form have a common divisor, then the problem of counting representations can be reformulated in terms of the form obtained by dividing all coefficients by this common divisor. To this end a form (**I**.14) is called primitive if $\gcd(a,b,c) = 1$, and from now on we will assume that all the forms are of this kind.

The second simplification is more subtle: if we consider a linear transformation of the plane $x = \alpha X + \beta Y$, $y = \gamma X + \delta Y$ inducing a bijection of $\mathbb{Z}^2$ into itself, then the quadratic form $Q(X,Y)$ also represents every integer exactly the same number of times as $Q(x,y)$ does, and hence for all purposes we may identify both forms. We say then that the forms are *equivalent*. It is an easy exercise to check that such transformations are the ones given by those matrices $\left(\begin{smallmatrix}\alpha & \beta\\ \gamma & \delta\end{smallmatrix}\right)$ with integer coefficients and determinant ± 1, and that when composed with quadratic forms they preserve the quantity $d = b^2 - 4ac$, called the *discriminant* of the form. Indeed, this second fact is evident from the matrix representation $Q(\vec{x}) = \vec{x}^t A \vec{x}$, as the change of variables may we written $\vec{x} = M\vec{X}$ for M with determinant ± 1 and thus invertible over $\mathbb{Z}$. Note from the definition of discriminant d that it is always an integer congruent to either 0 or 1 modulo 4, and reciprocally any such integer is the discriminant of either the primitive form $x^2 - (d/4)y^2$ or $x^2 + xy + (d-1)y^2/4$.

In some treatises quadratic forms are defined instead as a *rank two lattice* $\Lambda \subset \mathbb{R}^2$ endowed with a *quadratic function* $Q : \Lambda \to \mathbb{R}$. In this terminology, a rank n lattice refers to a discrete additive subgroup of $\mathbb{R}^n$ isomorphic to $\mathbb{Z}^n$, and a quadratic function is a function satisfying the axioms: *i)* $Q(ax) = a^2 Q(x)$ for any $a \in \mathbb{Z}$ and $x \in \Lambda$ and *ii)* the function $Q(x+y) - Q(x) - Q(y)$ is a bilinear form. Such an approach can be found, for example, in [**85**] and has the advantage of being basis-independent. After specifying an ordered basis of Λ as abelian group the function Q may then be identified with a concrete homogeneous polynomial of degree two evaluated at $\mathbb{Z}^2$. Different bases of the same lattice produce equivalent forms; and viceversa, equivalent forms may always be obtained from the same quadratic function by different bases of the same lattice. Sometimes the lattice also carries

extra structure which is not obvious from the quadratic form. One example of this is provided by the lattice of *Gaussian integers*

$$\mathbb{Z}[i] = \{a + bi \,:\, a, b \in \mathbb{Z}\} \subset \mathbb{C} \approx \mathbb{R}^2,$$

endowed by the quadratic function $Q(z) = |z|^2$. The classical identity $|z| \cdot |z'| = |zz'|$ satisfied by the modulus of complex numbers readily translates to the composition law (**I**.15) given above.[11]

We can generalize the composition law to other forms in a similar way by considering appropriate lattices obtained from algebraic number theory. We recall some elementary notions. Given a finite extension $\mathbb{K}$ of $\mathbb{Q}$ we may find inside the ring of *algebraic integers* (or number ring) $\mathcal{O}$ of $\mathbb{K}$, consisting of those elements of $\mathbb{K}$ whose monic minimal polynomial has integer coefficients. In many aspects $\mathcal{O}$ plays a role analogous to $\mathbb{Z}$, but often lacks unique factorization. This is not a big drawback because unique factorization is recovered at the ideal level: every ideal of $\mathcal{O}$ factors in an essentially unique way into prime ideals.[12] Ideals of $\mathcal{O}$ are also isomorphic, as abelian groups, to $\mathbb{Z}^n$ where n is the degree of the extension $\mathbb{K}/\mathbb{Q}$, and in fact they can be embedded in a more or less canonical way into $\mathbb{R}^n$ as rank n lattices. Since we are interested only in binary forms it makes sense to restrict our discussion to the case of extensions of degree two. These are always of the form $\mathbb{K} = \mathbb{Q}(\sqrt{D})$ for some square-free integer $D \neq 0, 1$. The norm $N : \mathbb{K} \to \mathbb{Q}$, defined as the product of all the Galois conjugates, can be seen to be a quadratic function when restricted to any ideal I of $\mathcal{O}$. To be more explicit: the general element of $\mathcal{O}$ can always be written in the form $a+b\sqrt{D}$ and $N(a+b\sqrt{D}) = a^2 - Db^2$. The canonical embedding into $\mathbb{R}^2$ when $D < 0$ is just the usual identification $\mathbb{C} \approx \mathbb{R}^2$, while when $D > 0$ it is given by $a + b\sqrt{D} \mapsto (a + b\sqrt{D}, a - b\sqrt{D}) \in \mathbb{R}^2$. Hence the associated quadratic function corresponds to the square of the Euclidean norm in the first case, and to the square of the "singular norm" $(x, y) \mapsto \sqrt{xy}$ in the second case.

After we fix a basis of the ideal I as an abelian group, let us say $\alpha, \beta \in I$, the norm function on I can be identified with the homogeneous polynomial

$$Q(x, y) = N(\alpha x + \beta y) = N(\alpha)x^2 + (\alpha'\beta + \alpha\beta')xy + N(\beta)y^2$$

where α' and β' are, respectively, the Galois conjugates of α and β. All three coefficients lie in $\mathcal{O} \cap \mathbb{Q} = \mathbb{Z}$, and therefore Q has integer coefficients, but it need not be primitive. In fact the gcd of all three coefficients can be seen to equal the index of I in $\mathcal{O}$, called the *norm of the ideal* and denoted by $N(I)$. In this fashion we can construct a primitive quadratic form $N(I)^{-1}Q(x, y)$ for every choice of an ideal and a basis of such ideal. The forms obtained in this way always have discriminant $d = D$ if D is congruent to 1 modulo 4 and $d = 4D$ otherwise, quantity which receives the

[11] The second statement in the composition theorem is more subtle. It follows from the fact that in $\mathbb{Z}[i]$ every element factorizes into prime elements, whose norm are either a prime number (if the element is not in $\mathbb{Z}$) or the square of a prime number.

[12] This is actually the reason ideals bear that name: Dedekind devised them as "ideal" factors, which would be required for the fundamental theorem of arithmetic to hold in number rings. Elements of the ring $\mathcal{O}$ can be identified, up to units, with principal ideals, and therefore all the non-principal ideals constitute "missing factors" from the viewpoint of elements of $\mathcal{O}$. An enlightening example due to Hilbert where something analogous happens is the multiplicative set of all integers congruent to 1 modulo 4. In this set we have $693 = 9 \times 77 = 21 \times 33$, factorizations which seem irreconcilable. After adding the missing factors $3, 7, 11$ however the problem disappears, as then both factorizations reveal to be the same with the factors grouped in two different ways: $(3 \times 3) \times (7 \times 11)$ and $(3 \times 7) \times (3 \times 11)$.

name *fundamental discriminant* of the field $\mathbb{Q}(\sqrt{D})$; and actually any form of this kind which is not negative definite can be constructed by this procedure. We are going to restrict our attention to these forms, as the negative definite ones can be replaced by their opposite, and the ones which have non-fundamental discriminant require considering ideals of full-rank subrings of $\mathcal{O}$ called *quadratic orders* which lie out of the scope of this survey.

When we replace the basis of I with another basis of the same ideal the above procedure of course produces an equivalent quadratic form, but this may also happen when we replace I by a different ideal. An example is given by the ideals I and aI, where a is any non-unit of $\mathcal{O}$ of positive norm.[13] Supposing there was no more redundancy we could quotient the set of all ideals of $\mathcal{O}$ by the relation $I \sim J$ if and only if there exist elements $a, b \in \mathcal{O}$ with norms of the same sign satisfying $aI = bJ$, to obtain a nice correspondence between classes of ideals and classes of quadratic forms. In general however the picture is more complicated than this, as there are ideals not related in any obvious way which give raise to equivalent forms.[14] The natural way to fix this turns out to be to consider a finer notion of equivalence between quadratic forms: two quadratic forms are said to be *properly equivalent* if they are related by a linear transformation with integer coefficients and discriminant $+1$, thereby excluding the ones with discriminant -1. In other words, we require the linear transformation to fix an *orientation* of the plane. This also forces us to consider only those bases of ideals which are *positively oriented*, as determined by the embedding into $\mathbb{R}^2$ described above. With these amendments we have the following correspondence:

Theorem (Correspondence between ideals and forms). *Let d a fundamental discriminant and $\mathcal{O}$ the number ring of $\mathbb{Q}(\sqrt{d})$. Then:*

(*i*) *Any primitive quadratic form of discriminant d which is not negative definite can be obtained via the aforementioned procedure from some positively oriented basis of some ideal of $\mathcal{O}$; and viceversa all quadratic forms obtained in this way are of this kind.*

(*ii*) *Let I and J be two ideals of $\mathcal{O}$ and fix two positively oriented bases of them. The forms thus obtained are properly equivalent if and only if $I \sim J$.*

Choose now bases α_1, α_2 of I, β_1, β_2 of J and γ_1, γ_2 of IJ, the product ideal. There are integers a_{ijk} satisfying $\alpha_i \beta_j = a_{ij1}\gamma_1 + a_{ij2}\gamma_2$, and therefore

$$(x_1\alpha_1 + x_2\alpha_2)(y_1\beta_1 + y_2\beta_2) = \Big(\sum_{i,j} a_{ij1}x_iy_j\Big)\gamma_1 + \Big(\sum_{i,j} a_{ij2}x_iy_j\Big)\gamma_2.$$

Taking norms, dividing by the norm of IJ, and using that the norm is multiplicative for both elements and ideals, we obtain the composition law

$$Q_I(x_1, x_2)Q_J(y_1, y_2) = Q_{IJ}\Big(\sum_{i,j} a_{ij1}x_iy_j, \sum_{i,j} a_{ij2}x_iy_j\Big),$$

[13] Choose bases α, β and $a\alpha, a\beta \in aI$ and use $N(aI) = |N(a)|N(I)$.

[14] An example can be sketched as follows: consider the forms $3x^2+2xy+5y^2$ and $3x^2-2xy+5y^2$, which are both of fundamental discriminant -56 and equivalent. They arise from the ideals I and J, generated (as abelian groups) by 3 and $1+\sqrt{-14}$, and by 3 and $-1+\sqrt{-14}$, respectively. In both lattices the shortest nonzero vectors are ± 3, hence if $aI = bJ$ we must have either $a = b$ or $a = -b$, and in both cases $I = J$, a contradiction.

where Q_I is the quadratic form associated to the basis α_1, α_2 of I, and so on. If we forget the points where we are evaluating the forms this also provides a product law $[Q_I] \cdot [Q_J] = [Q_{IJ}]$ between classes of properly equivalent quadratic forms of discriminant d. A very surprising and absolutely non-trivial fact is that, endowed with the product thus defined, the set of such equivalency classes is a finite abelian group, called the *narrow class group* of $\mathbb{Q}(\sqrt{d})$. This group, of course, can also be defined directly from the viewpoint of $\mathcal{O}$ by endowing the set of classes of ideals with the ideal product.

Far more complicated examples of composition laws are obtained when the narrow class group is not trivial. For example for $d = -20$ the narrow class group is isomorphic to $\mathbb{Z}/2\mathbb{Z}$, its elements being given by the equivalence classes of the quadratic forms $Q_1(x, y) = x^2+5y^2$ and $Q_2(x, y) = 2x^2+2xy+3y^2$. The composition laws then are

$$\begin{aligned} Q_1(x, y)\, Q_1(x', y') &= Q_1(xx' - 5yy', x'y + xy') \\ Q_1(x, y)\, Q_2(x', y') &= Q_2(xx' - x'y - 3yy', xy' + 2x'y + yy') \\ Q_2(x, y)\, Q_2(x', y') &= Q_1(2xx' + xy' + x'y - 2yy', xy' + x'y + yy'). \end{aligned}$$

If one allows quadratic forms to be also related by matrices of determinant -1 (equivalence) then not only the correspondence theorem breaks down but it is also impossible to define a meaningful group law, or even a well-defined product, on the resulting set of classes. Gauss noticed this himself and introduced the notion of proper equivalence in his Disquisitiones. He then was able to define the product law and work out all the details, including the fact that the set of classes constitutes a finite group. This was remarkably done without the modern algebraic machinery (not even the definition of group!), relying instead on a convoluted casuistic and elemental number theory manipulations, making the proof a real *tour de force*.

On the other hand, if we relax the equivalence relation of ideals by not requiring the factors to have norms of the same sign then we obtain another finite abelian group, called the *class group* of $\mathbb{Q}(\sqrt{d})$. This object is arguably more natural from the algebraic point of view than its narrow version, but very often both groups coincide (and when they do not the class group is always a quotient of the narrow class group by a subgroup of order two). The order of the class group, called the *class number* and denoted $h(d)$, is an important but poorly-understood arithmetic function.[15] In the case $d < 0$, where both class groups have the same size, $h(d)$ is also given by the number of elements of a complete set of representatives $Q_1, \ldots, Q_{h(d)}$ of the proper equivalence classes of positive definite forms of discriminant d.[16] Understanding how many times each integer n is represented by each of the forms Q_i is in general a very difficult problem, but there is a nice formula (also due to Gauss) giving the total number $R(n)$ of representations of the integer n provided by all the forms

[15]Many questions about the growth of $h(d)$ are still open. For example, when $d < 0$ we have $h(d) = 1$ only for $d = -3, -4, -7, -8, -11, -19, -43, -67$ and -163 (the Stark-Heegner theorem), but the question of how many times $h(d) = 1$ is still open for $d > 0$. Gauss conjectured this should happen infinitely often.

[16]For fundamental $d > 0$ the class number is either the number or half the number of proper equivalence classes of forms. If d is not fundamental, the class number may also be defined in a similar way, but considering only primitive forms. In general, the values of $h(d)$ for non-fundamental d are not that interesting, as they are easily related to sums of $h(d)$ for fundamental d (theorem 2 of chapter XIII of [**22**]).

$Q_1, \dots, Q_{h(d)}$ at once (see §12.4 of [**56**]):

$$R(n) = w \sum_{m|n} \left(\frac{d}{m}\right), \qquad \text{where} \quad w = \begin{cases} 6 & \text{if } d = -3, \\ 4 & \text{if } d = -4, \\ 2 & \text{otherwise.} \end{cases} \tag{**I**.16}$$

On the right hand side $(\frac{d}{m})$ stands for the Jacobi symbol (see chapter 5 of [**23**]). Note this identity generalizes the genus theorem given above.[17] The same formula also holds when $d > 0$ with $w = 1$, but then it only counts a special kind of representations (primary representations) which are finite in number (see chapter 6 of [**23**]). To avoid these technical details we assume $d < 0$ from now on.

The function $R(n)$ as defined is very irregular, but in average its behavior is quite smooth. In fact, assuming each Q_i contributes more or less the same, the average of $R(n)$ must be proportional to $h(d)$. Dirichlet succeeded in using this idea to obtain a formula for computing the class number, the *Dirichlet class number formula*. We are going to derive this formula following the exposition by Davenport (chapter 6 of [**23**]). We start by considering, for convenience, the average of $R(n)$ only over those values of n coprime to d:

$$S(n) = \frac{1}{n} \sum_{\substack{m \le n \\ \gcd(m,d)=1}} R(m).$$

The idea is to expand this sum by substituting (**I**.16) and then using Dirichlet's hyperbola method to estimate the double sum:

$$\begin{aligned} nS(n) &= w \sum_{\substack{m_1 m_2 \le n \\ \gcd(m_1 m_2, d)=1}} \left(\frac{d}{m_1}\right) \\ &= \sum_{m_1 \le \sqrt{n}} \left(\frac{d}{m_1}\right) \sum_{\substack{m_2 \le n/m_1 \\ \gcd(m_2,d)=1}} 1 + \sum_{\substack{m_2 < \sqrt{n} \\ \gcd(m_2,d)=1}} \sum_{\sqrt{n} < m_1 \le n/m_2} \left(\frac{d}{m_1}\right). \end{aligned}$$

The first double sum is approximately $n\phi(|d|)|d|^{-1} \sum_{m \le n} m^{-1} (\frac{d}{m})$, where ϕ is Euler's totient function, while the second double sum must be small because of the cancellation provided by the character $\chi_d(\cdot) = (\frac{d}{\cdot})$. Therefore

$$\lim_{n \to \infty} S(n) = w \frac{\phi(|d|)}{|d|} \sum_{m \ge 1} \frac{1}{m} \left(\frac{d}{m}\right) = w \frac{\phi(|d|)}{|d|} L(1, \chi_d).$$

On the other hand, if for each quadratic form Q we define $r_Q(n)$ as the number of representations of n by Q, then $R(n) = \sum_i r_{Q_i}(n)$ and

$$nS(n) = \sum_{i=1}^{h(d)} \sum_{\substack{m \le n \\ \gcd(m,d)=1}} r_{Q_i}(m).$$

Let us forget for a second the coprimality condition. The sum $\sum_{m \le n} r_Q(m)$ can be interpreted as the number of points with integer coordinates lying in the region of

[17] There is a further generalization due to Gauss. He noticed that sometimes each of the forms Q_i represent disjoint sets of integers, and therefore $R(n)$ coincides with the number of representations coming from a specific Q_i. This is the *genus* theory: each genus consists of forms which essentially represent the same integers. When a genus contains more than one proper equivalence class very little can be said about the representation problem (*cf.* §XIII.3 of [**22**]).

the xy plane determined by $Q(x,y) \leq n$. This quantity must be well approximated by the volume of the region, as the points are well-distributed and the region has a "simple" shape. The following rigorous argument of this fact is attributed to Gauss: let us draw a square of side-length unity around each point with integer coordinates. Any region formed by an union of these squares contains exactly as many points with integer coordinates as area covers. Consider now the region Ω composed of all those squares whose center lies inside the ellipse $Q(x,y) \leq n$. By the previous remark the area of Ω is exactly $\sum_{m \leq n} r_Q(m)$. But the area of Ω is almost equal to that of the ellipse: to make them coincide we just have to add or remove disjoint pieces of those squares which intersect the contour $Q(x,y) = n$. Therefore

$$\sum_{m \leq n} r_Q(m) = \mathrm{Vol}\{Q(x,y) \leq n\} + O\big(\,\mathrm{diam}\{Q(x,y) \leq n\}\big).$$

Some elemental analysis shows that the area of the ellipse equals $2\pi n|d|^{-1/2}$, while the error term is of order $O(\sqrt{n})$ for fixed d. Therefore

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{h(d)} \sum_{m \leq n} r_{Q_i}(m) = \frac{2\pi h(d)}{|d|^{1/2}}.$$

If we restore the coprimality condition the same result is still true with an extra factor $\phi(|d|)|d|^{-1}$. To see this note that if $Q(n_1,n_2) = m$ then the residues of n_1 and n_2 modulo d determine that of m. Hence our sum counts points (x,y) in the ellipse $Q(x,y) \leq n$ whose coordinates are integers that modulo d lie in a certain subset of $\mathbb{Z}/d\mathbb{Z} \times \mathbb{Z}/d\mathbb{Z}$, which is easily shown to be of size $\phi(|d|)|d|$. Considering squares this time of side-length $|d|$ instead of 1 we arrive, by the same argument, to

$$\lim_{n\to\infty} S(n) = \frac{2\pi h(d)\phi(|d|)}{|d|^{3/2}}.$$

Putting together the two expressions we have for the limit of $S(n)$ we obtain the Dirichlet class number formula:

$$h(d) = \frac{w}{2\pi}|d|^{1/2}L(1,\chi_d). \tag{I.17}$$

This result has deep implications; for example it readily shows that $L(1,\chi)$ does not vanish when χ is a non-trivial real caracter, which is an essential ingredient of Dirichlet's theorem on the infinitude of primes in an arithmetic progression (see §1 of [**23**]). It can also be used as an efficient way of computing the class number $h(d)$, by either approximating $L(1,\chi_d)$ or by employing Gauss sums to express the value of the L-function as a finite sum (see equations (17) and (18) of §6 of [**23**]).

There is evidence that Gauss already knew this formula almost forty years before it was published by Dirichlet, and it is in this regard he came up with the aforementioned argument used to count points with integer coordinates ("lattice points") in ellipses.[18] In the simplest case, $d = -4$, we have $h(d) = 1$ (as shown, for example, by the class number formula and the arctangent Taylor series), and the only proper equivalence class of forms is represented by $Q(x,y) = x^2 + y^2$. Since

[18]For negative discriminants (as stated here) this formula can be found in Gauss' article [**34**], published in 1837, two years before Dirichlet published his work. Gauss motto "pauca sed matura" (few, but ripe) would often led him to publish his work many years after coming up with an idea. In this article Gauss counts points with integer coordinates in circles and ellipses by slicing them into small squares, essentially as described above.

$R(n) = r_Q(n)$, its sum $\mathcal{N}(R) = \sum_{n \leq R^2} r_Q(n)$ counts the number of lattice points inside the circle $x^2 + y^2 \leq R^2$. Gauss' argument shows

$$\mathcal{N}(R) = \pi R^2 + O(R).$$

If one estimates numerically the error term $\mathcal{N}(R) - \pi R^2$, they will notice that it seems to become a lot smaller than this result suggests. For example, for $R = 100$ we have $\mathcal{N}(100) = 31417$, while $\pi 10^4 = 31415.92...$; the error term is smaller than $1 = 0.01R$. For higher values of R this trend goes on. A sharper estimation of the error term however would have to wait until 1906, when Sierpiński [**89**] proved, using ideas from Voronoï, that

$$\mathcal{N}(R) = \pi R^2 + O(R^{2/3}). \tag{**I**.18}$$

This is surprising as no naive geometrical intuition shows us why the error term should have power-savings over R. The problem of determining the infimum of the values of α for which $\mathcal{N}(R) = \pi R^2 + O(R^\alpha)$ holds is still open, and has become known as *Gauss' circle problem*. The sharpest result at the time of writing this dissertation is due to Bourgain and Watt [**11**], who using a method developed by Huxley [**59**] have shown that for any $\alpha > 517/824 \approx 0.627$ the estimate above is true.[19] On the other hand, in 1915 Hardy and Landau [**41, 73**] proved independently that $\mathcal{N}(R) = \pi R^2 + O(\sqrt{R})$ cannot hold, and Hardy went on to conjecture that the error term should be $O(R^{1/2+\epsilon})$ for any $\epsilon > 0$; *i.e.*, the aforementioned infimum should be $1/2$.[20]

All modern techniques used to obtain non-trivial estimations for Gauss' circle problem make use, as a first step, of the Fourier transform via the *Poisson summation formula*.[21] This translates the problem of bounding the error term into one of bounding exponential sums. We are going to sketch a modern proof of Sierpiński's result (**I**.18) using these tools, to illustrate the underlying ideas.

We begin by considering χ_R the characteristic function of the circle of radius R centered at the origin. Applying the Poisson summation formula,

$$\mathcal{N}(R) = \sum_{\vec{n} \in \mathbb{Z}^2} \chi_R(\vec{n}) = \pi R^2 + \sum_{\vec{0} \neq \vec{n} \in \mathbb{Z}^2} \hat{\chi}_R(\vec{n}).$$

[19] Voronoï's original idea consists in approximating the circle by a convex inscribed polygon, whose sides have slopes which are rational numbers p/q of bounded p and q. Huxley further refined this method by replacing the straight edges by pieces of conics, idea originally developed by Bombieri and Iwaniec [**10**] to study the size of $\zeta(1/2 + it)$. The method then becomes very analytic and resembles the Hardy-Littlewood method, as the main contribution comes from those pieces of the curve with slopes really close to those of the edges of the Voronoï-Sierpiński polygon. See Huxley's book [**58**] or the survey [**57**] for further detail. This method in the literature usually receives the name of the *discrete Hardy-Littlewood method*.

[20] Some authors refer to the following heuristics: let A_i be the area of the circle lying inside one of the one by one squares intersecting the boundary of the circle, and let $P_i = 1$ if the center of the square lies inside the circle and $P_i = 0$ otherwise. Assuming the quantities $A_i - P_i$ behave like independent random variables with zero mean, and since there are about R of them, the central limit theorem suggests the error of the circle problem to be bounded by $R^{1/2+\epsilon}$. Why the curvature of the circle should imply the independence is not clear to me. This argument also seems to break in higher dimensions.

[21] This result establishes the equality $\sum f(n) = \sum \hat{f}(n)$, essentially as long as both sums converge, where $\hat{f}$ is the Fourier transform of f and n runs over $\mathbb{Z}$ in both sums. An analogous version holds if $\mathbb{Z}$ replaced by $\mathbb{Z}^n$. See the appendix for a proof.

An "explicit" expression for the Fourier transform of χ_R can be given in terms of Bessel's function J_1:

$$\hat{\chi}_R(\vec{\xi}) = R\frac{J_1\big(2\pi R\|\vec{\xi}\|\big)}{\|\vec{\xi}\|}.$$

Substituting above and performing the change of variables $\|\vec{n}\| = \sqrt{n}$ we arrive to the identity

$$\mathcal{N}(R) = \pi R^2 + R\sum_{n\geq 1} r_2(n)\frac{J_1\big(2\pi R\sqrt{n}\big)}{\sqrt{n}} \tag{I.19}$$

where $r_2(n) = r_Q(n)$ denotes the number of different ways of expressing n as a sum of two squares.

At this point we have to admit that the argument we have given to obtain (**I**.19) is fallacious, as the lack of regularity of the function χ_R translates to a very poor decay for its Fourier transform, making the convergence of the series $\sum \hat{\chi}_R(\vec{n})$ a very subtle matter. Formula (**I**.19) is actually true when restricted to values of R which are not the square root of an integer, as shown by Hardy [**43**], but the proof is by no means this simple. Nevertheless we can still rigorously apply the Poisson summation formula if we first mollify χ_R, obtaining a weaker version of (**I**.19) which is enough for our purposes. With this objective in mind, we pick a radial bump function $\eta \in \mathcal{C}^\infty(\mathbb{R}^2)$ satisfying for some $h = h(R) \leq 1$ to be chosen later,

$$\eta \geq 0, \qquad \int \eta = 1 \qquad \text{and} \qquad \operatorname{supp}\eta \subset B(0,h).$$

Note that the difference between $\mathcal{N}(R)$ and $\sum \chi_R * \eta(\vec{n})$ can always be bounded by the number of points of $\mathbb{Z}^2$ lying in the annulus of radii $R-h$ and $R+h$, precisely given by the sum

$$\sum_{(R-h)^2\leq m\leq (R+h)^2} r_2(m).$$

We are going to employ that for any $\epsilon > 0$ the bound $r_2(n) \ll n^\epsilon$ holds. To see this is true note first that the divisor function σ_0, counting the number of divisors of an integer, does satisfy $\sigma_0(n) \leq n^\epsilon$ for n big enough, since both sides are multiplicative and the result is trivial for prime powers. Also by (**I**.16) we have $r_2(n) \leq 4\sigma_0(n)$, and hence $r_2(n) \ll n^\epsilon$ as claimed. Therefore,

$$\begin{aligned}\mathcal{N}(R) + O\big(hR^{1+\epsilon}\big) &= \sum_{\vec{n}\in\mathbb{Z}^2} \chi_R * \eta(\vec{n}) = \pi R^2 + \sum_{\vec{0}\neq\vec{n}\in\mathbb{Z}^2} \hat{\chi}_R(\vec{n})\cdot\hat{\eta}(\vec{n}) \\ &= \pi R^2 + R\sum_{n\geq 1} r_2(n)\hat{\eta}\big(\sqrt{n}\big)\frac{J_1\big(2\pi R\sqrt{n}\big)}{\sqrt{n}},\end{aligned}$$

where we have written $\hat{\eta}(\sqrt{n})$ instead of $\hat{\eta}(\sqrt{n},0)$ for the sake of clarity. Note that the smoothness of η implies that $\hat{\eta}$ is of fast decay, forcing the sum to converge. In fact we may choose η satisfying that almost all the mass of $\hat{\eta}$ lies in $B(0,h^{-1-\epsilon})$, allowing us to truncate the sum up to a small error term $O(h^{-\epsilon}R^\epsilon)$. Using widely known asymptotics for J_1 (*cf.* chapter VII of [**94**]), namely

$$J_1(x) \sim \sqrt{\frac{2}{\pi x}}\cos\left(x - \frac{\pi}{4}\right) \ll \frac{1}{\sqrt{x}}, \tag{I.20}$$

and the aforementioned bound $r_2(n) \ll n^\epsilon$ we obtain

$$\mathcal{N}(R) + O(hR^{1+\epsilon}) = \pi R^2 + O\left(R^{1/2}h^{-5\epsilon/2} \sum_{1\le n\le h^{-2-2\epsilon}} \frac{1}{n^{3/4}}\right) + O(h^{-\epsilon}R^\epsilon)$$

$$= \pi R^2 + O\Big(h^{-\frac{1}{2}-3\epsilon}R^{1/2}\Big). \tag{I.21}$$

Choosing now $h = R^{-1/3}$ we conclude $\mathcal{N}(R) = \pi R^2 + O(R^{2/3+\epsilon})$ for any $\epsilon > 0$.

The same proof may be adapted with minimal changes to remove the extra ϵ in the following way: taking $\eta(\vec{x}) = h^{-1}\eta_0(\vec{x}/h)$ for a fixed η_0 not depending on h, we obtain the uniform bound $\hat{\eta}(x) \ll \min\big(1, (xh)^{-1}\big)$. Summing by parts and using the estimation obtained by Gauss for the circle,

$$\sum_{n\ge 1} \frac{r_2(n)}{n^{3/4}} |\hat{\eta}(\sqrt{n})| \ll \sum_{1\le n\le h^{-2}} \frac{r_2(n)}{n^{3/4}} + \sum_{n>h^{-2}} \frac{r_2(n)}{hn^{5/4}}$$
$$= 3\pi(h^{-2})^{\frac{1}{4}} + 5\pi h^{-1}(h^{-2})^{-\frac{1}{4}} + O(1).$$

This upper bound suffices to remove the ϵ on the right hand side of (**I**.21). To remove the other one, note that the inequalities

$$\sum_{\vec{n}\in\mathbb{Z}^2} \chi_{R-h} * \eta(\vec{n}) \le \mathcal{N}(R) \le \sum_{\vec{n}\in\mathbb{Z}^2} \chi_{R+h} * \eta(\vec{n})$$

imply, by the same argument leading to (**I**.21),

$$\begin{cases} \mathcal{N}(R) \le \pi(R+h)^2 + O\big(h^{-\frac{1}{2}}(R+h)^{\frac{1}{2}}\big), \\ \mathcal{N}(R) \ge \pi(R-h)^2 + O\big(h^{-\frac{1}{2}}(R-h)^{\frac{1}{2}}\big). \end{cases}$$

Choosing, again, $h = R^{-1/3}$, we conclude $\mathcal{N}(R) = \pi R^2 + O(R^{2/3})$.

To go beyond Sierpiński's exponent 2/3 one has to take advantage of the cancellation provided by the sign of the cosine in (**I**.20); *i.e.*, one essentially has to find non-trivial bounds for the exponential sum

$$\sum_{n\ge 1} \hat{\eta}(\sqrt{n}) \frac{r_2(n)}{n^{3/4}} e(R\sqrt{n}).$$

Here $e(x)$ stands for $e^{2\pi i x}$. Note we may assume $\hat{\eta} > 0$ by appropriately choosing η, for example as a convolution of a function with itself. Summing by parts we can then remove the smooth factor $\hat{\eta}(\sqrt{n})n^{-3/4}$, reducing the problem to that of bounding the exponential sum $\sum_{m\le n} r_2(m)e(R\sqrt{m})$ in terms of R and n. To avoid the highly irregular factor r_2 it is convenient to take a step backwards and rewrite the sum as $\sum_{m_1^2+m_2^2\le n} e\Big(R\sqrt{m_1^2+m_2^2}\Big)$.

Van der Corput devised a general method to estimate exponential sums of the form $\sum e(\phi(m))$ for a smooth phase function $\phi : \mathbb{R} \to \mathbb{R}$, consisting in two processes which transform the sum, with the objective of arriving to an exponential sum of shorter length. If this is achieved then one may trivially estimate the resulting sum by the number of summands to obtain a non-trivial bound for the original sum. The two procedures can be roughly described as either squaring the modulus of the sum or applying Poisson's summation formula, and the bounds obtained in this way are referred to as *van der Corput estimates*. We will devote section §4.4 to explain them in more detail. Even the simplest van der Corput estimates suffice to obtain non-trivial results for Gauss' circle problem beyond Sierpiński's 2/3. Despite

this the method has its limitations, and for this particular problem the proof of the aforementioned result due to Bourgain, Watt and Huxley is more closely related to the original ideas of Voronoï and Sierpiński than to those of van der Corput.

Nowadays Gauss' circle problem is the most paradigmatic of a loosely defined family of related problems, receiving the name of *lattice point counting problems*. The objective is always estimating the number of points in a lattice (without loss of generality $\mathbb{Z}^d \subset \mathbb{R}^d$) that lie in a certain region, depending on one or more parameters. For example, the sum $\sum_{m\leq n} \sigma_0(m)$, essentially the average of the divisor function, can also be interpreted as counting the number of points with integer coordinates lying in the two-dimensional hyperbolic region

$$\{xy \leq n,\, 1 \leq x, y \leq n\}.$$

The volume of this region is $n \log n - n + 1$, while the perimeter $O(n)$. Gauss' argument therefore shows that the average of the divisor function over the first n integers is asymptotically $\log n$. This problem is usually regarded as *Dirichlet's divisor problem*. As with the circle problem the error term is actually smaller, and in fact these two problems are closely related to each other [**11**]. Other examples of lattice point counting problems arising from number theory include the average of the class number [**18**] or the equidistribution of rational points on the unit sphere [**25**]. Even some Diophantine approximation problems (such as well-approximability) can be rephrased as determining if there are infinitely many points with integer coordinates in certain regions.

The same techniques we have sketched so far can also be applied to many other similar problems. In particular, to that of counting points with integer coordinates lying inside a fixed d-dimensional convex body, after being dilated by a factor $R > 0$, as long as its boundary is a smooth manifold and has positive Gaussian curvature. The restriction on the curvature is necessary, as shown for example by the square centered at the origin, for which the error term is infinitely often as big as the perimeter.

Once the lattice point counting problem has been reformulated as bounding the corresponding exponential sum, obtaining sharp estimates is usually a very difficult task. To give a sense of the state of the art, let $\mathcal{N}(R)$ denote the number of points with integer coordinates lying inside the convex body after being dilated by the factor $R > 0$, V its volume for $R = 1$, d the dimension of the ambient space, and assume the asymptotic $\mathcal{N}(R) = VR^d + O(R^{\alpha+\epsilon})$ holds for any $\epsilon > 0$. For the plane, $d = 2$, the best known result has been obtained by Huxley [**59**] using a refinement of the original ideas of Voronoï and Sierpiński, yielding $\alpha = 131/208 \approx 0.63$.[22] When $d \geq 3$ the best known result is due to Guo [**39**], who used a bidimensional version of the van der Corput method to obtain $\alpha = d-2+r(d)$, where $r(d) = 73/158 \approx 0.462$ for $d = 3$ and $r(d) = (d^2 + 3d + 8)/(d^3 + d^2 + 5d + 4)$ for $d \geq 4$. These results are still quite far from the conjectured $\alpha = 1/2$ for $d = 2$ (same as for the circle) and $\alpha = d - 2$ for $d \geq 3$. See chapter 4 for more information.

If one adds extra hypotheses on the convex body, or restricts it to very particular shapes, sometimes the corresponding exponential sum is better understood and

[22] It might be possible to translate the result of Bourgain and Watt [**11**] for the circle, *i.e.* $\alpha = 517/824$, to any convex body in the plane satisfying the above hypothesis on the curvature. This is because we do not know how to take advantage of the fact that the circle is a very special region, and the techniques are rather generic.

therefore one may obtain better bounds. One of these special cases is provided by the parabolic region[23]

$$\{|y| \leq R - x^2/R\}.$$

Popov [**81**] noticed that the corresponding exponential sum is quadratic, essentially $\sum_{|n|\leq N} e(n^2 x)$, and therefore of the kind studied by Hardy and Littlewood. Using these ideas he was able to obtain the sharp exponent $\alpha = 1/2$, precisely as conjectured for the circle. We will give a simplified version of his proof in chapter 5.

Recall that we mentioned these sums may be estimated by considering them a truncated version of Jacobi's theta function, and then relating their size to the size of $|\theta(x+iy)|$ for $y \approx 0$. This, in turn, can be bounded by using the functional equation that θ satisfies for being a modular form. The advantage of this proof is that it generalizes well to other modular forms, and in particular to the powers θ^k, allowing us to obtain very sharp bounds for the k-dimensional exponential sums

$$\sum_{n_1^2+\cdots n_k^2 \leq N} e\big((n_1^2+\cdots n_k^2)x\big) = \sum_{n\leq N} r_k(n) e(nx),$$

where the function $r_k(n)$ counts the number of different ways of writing n as a sum of k squares. These, for $k = d-1$, correspond to the lattice point counting problem associated to the d-dimensional paraboloid

$$\left\{ |x_d| \leq R - \frac{1}{R} \sum_{i=1}^{d-1} x_i^2 \right\}.$$

In [**20**] Chamizo and the author used these ideas to obtain the conjectured exponent $\alpha = d-2$ for this family of paraboloids. The result is interesting for $d = 3$ because, as far as the authors know, it constitutes the first non-trivial example of a three-dimensional convex body for which the conjecture has been proved. In fact, the difficult step of the proof is precisely this special case, and then summation by parts suffices to generalize the bound to any $d > 3$. The case $d = 3$ is also closely connnected to binary quadratic forms, as the paraboloid is a dilation of $\{|z| \leq 1-(x^2+y^2)\}$, and the exponential sum, a truncated version of $\theta^2(z) = \sum r_2(n) e^{\pi i n z}$. If one replaces x^2+y^2 by any other binary quadratic form $Q(x,y)$ with integer coefficients, then the exponential sum becomes a truncated version of $\theta_Q(z) = \sum r_Q(n) e^{\pi i n z}$, which turns out to be again a modular form called the *theta function associated to* Q. The proof still works, *mutatis mutandis*, providing sharp exponents for a wider family of "elliptic" paraboloids. This will be presented in chapter 5.

Although the results on parabolic regions have interest *per se*, Chamizo and the author were lead to them while trying to gain intuition on a different problem. The original objective was to generalize the main result of the article [**15**], which we describe in what follows. Consider a convex body in three dimensions whose boundary is a smooth surface with positive Gaussian curvature, containing the origin, and invariant by rotations around the z-axis. Denote by $f(r)$ the generatrix of the convex body, parametrized by the radius $r^2 = x^2+y^2$ (see figure **I**.4). If one assumes that the quotient $f'''(r)/r$ never vanishes,[24] then exploiting the extra symmetry it

[23] The boundary in this example is not smooth at two points, which we may regard as having "infinite" Gaussian curvature. This is a minor technical problem of limited importance which will be ignored for now.

[24] This should be understood in the following sense: $f'''(r)$ never vanishes for $0 < r < r_0$ and neither does $f^{(4)}(0) = \lim_{r\to 0^+} f'''(r)/r$. Note that f is a two-valued function and hence we must ask this to hold for both branches.

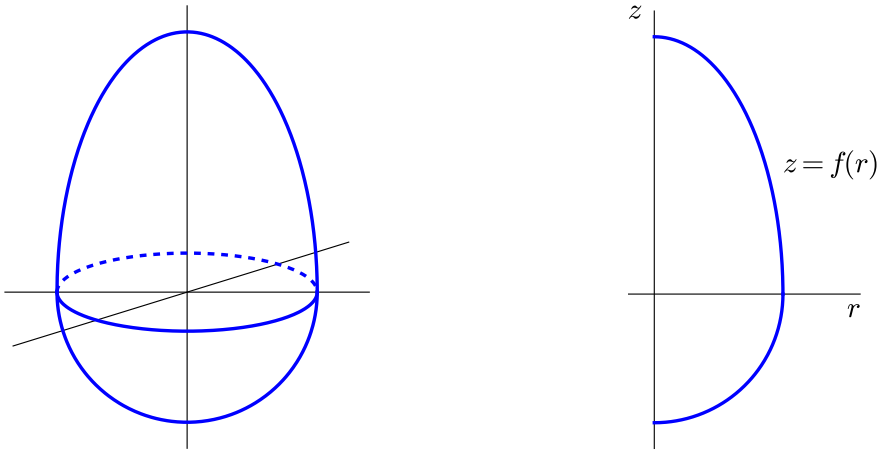


FIGURE I.4. On the left an example of smooth and convex revolution body. On the right its generatrix f parametrized by the radius r. Note f is a two-valued function.

is possible to go beyond Guo's result and obtain the exponent $\alpha = 11/8 = 1.375$. Improvements this big (0.087 over Guo's exponent 231/158) are rare, specially when dealing with exponential sums, but the nonvanishing condition —involving a third derivative— makes this result by Chamizo not completely satisfactory. This is so because heuristically only derivatives up to order two (with "geometrical" meaning) should matter to determine the size of the error term. We therefore proposed to try to weaken this hypothesis. A natural first step in this direction is to try to study the most pathological case: the case when f''' vanishes identically and therefore the condition fails at every point, resulting in a paraboloid. After the investigation it turned out that the techniques employed could not be readily translated to the original problem (which is no surprise as the paraboloid is a very special and arithmetic object) and instead we had to rely on a convoluted combination of van der Corput estimates, which at some point did include an arithmetic argument vaguely reminiscent of the one used for the paraboloid. In this way we succeeded in proving that the exponent $\alpha = 11/8$ still holds under the weaker hypothesis of asking all the zeros of $f'''(r)/r$ to be of *finite order*[25]. This, in particular, includes the case of f being real analytic. The theorem was published in [**21**], and will be presented in detail in chapter 6.

I.4. Outline of this document

This dissertation may be divided in two markedly different parts. Chapters 1, 2 and 3 focus on modular forms and their properties, while chapters 4, 5 and 6 are concerned with lattice point counting problems. These two parts are not completely independent, as chapter 5 depends upon some results obtained in §2.7.

The first two chapters can be thought as a very short course in classical holomorphic modular forms, for which we have assumed no background knowledge. All the material exposed there can be found in standard books (some of them cited

[25]The point x_0 is said to be a zero of finite order of the function g if $g(x_0) = 0$ but for some $n > 0$ we have $g^{(n)}(x_0) \neq 0$.

in the corresponding chapters), except for the contents of section §2.7. This section comprises two technical lemmas which are key to the results later exposed in chapters 3 and 5, and were originally part of the articles [**20, 80**].

Chapter 3 builds upon the first two chapters, presenting the contents of the article "On the regularity of fractional integrals of modular forms" [**80**], introduced in §**I**.2.

Chapter 4 briefly describes the state of the art in lattice point counting theory, and then introduces some widely used techniques. Again no background knowledge has been assumed, and the material can be found in many standard textbooks.

Finally, chapters 5 and 6 focus on the problems introduced in §**I**.3, corresponding to the articles "Lattice points in elliptic paraboloids" [**20**] and "Lattice points in bodies of revolution II" [**21**].

Each of the three chapters presenting research material (chapters 3, 5 and 6) includes a section named "Main results" where the original results are rigorously stated and compared with the existing literature prior to them. These chapters follow closely the content of the corresponding articles, although some parts are more carefully explained and some proofs of technical lemmas borrowed from other articles are included for convenience.

At the end of this dissertation the reader will find a short appendix containing some widely used tools in analytic number theory. These are results that the author found himself consulting again and again during the research.

CHAPTER 1

The modular group

In this chapter we introduce the modular group and some of its many arithmetic properties. This group will play a fundamental role in chapter 2 when defining modular forms. The presented results are chosen with the aim of helping the reader develop some intuition on the underlying theory, should this be their first encounter with the topic.

1.1. Lattices and the upper half-plane

We begin by providing some generalities about lattices. A lattice of rank n for us will be a discrete subgroup of $\mathbb{R}^n$ isomorphic to $\mathbb{Z}^n$. Discrete means that we may find a neighborhood around each point of the lattice containing only this point. Equivalently, it is discrete if and only if it intersects every compact subset of $\mathbb{R}^n$ at a finite number of points. A third equivalent condition: it is discrete if and only if it is generated (as an abelian group) by a basis of $\mathbb{R}^n$. A set of generators which is a basis of $\mathbb{R}^n$ is called a basis of the lattice, and will bear the adjective positively oriented if the linear map sending the canonical basis of $\mathbb{R}^n$ to the lattice basis has positive determinant. This classifies all bases of the lattice in positively or negatively oriented.

The group of matrices of size $n \times n$ with integer entries and determinant ± 1 acts transivitely and freely on the bases of the lattice. In other words, given any basis, the linear combinations of vectors dictated by the rows of the matrix always provide another basis, and as we change the matrix we obtain all possible bases once and only once. If we restrict to matrices of determinant $+1$ the same is true for the set of positively (or negatively) oriented bases, while matrices of determinant -1 invert the orientation of each basis. Matrices of determinant $\pm n$ with $n > 1$ provide all bases of sublattices of index n.

The group of all matrices of size $n \times n$ with integer entries and determinant $+1$ is called the special linear group of degree n over $\mathbb{Z}$, denoted $\mathrm{SL}_n(\mathbb{Z})$ or $\mathrm{SL}(n, \mathbb{Z})$. The only one of interest to us is the one of degree 2. In the case of lattices of rank two, a basis is positively oriented if and only if the angle measured from the first vector of the basis to the second lies in the interval $(0, \pi)$.

The fact that $\mathbb{R}^2$ may be identified with $\mathbb{C}$ gives us extra structure to play with. Given a lattice Λ and a nonzero complex number λ the lattice $\lambda\Lambda = \{\lambda l : l \in \Lambda\}$ is the result of letting a rotation followed by a homothety (both with respect to the origin) act on Λ. These two lattices have the same "shape", although dilated and tilted with respect to each other. The converse is also true: any two lattices which may be made to coincide by a rigid motion fixing the origin must be related in this way. Motivated by this, let $\mathcal{L}$ denote the set of all lattices of rank two and define the map $\phi : \mathcal{L} \to \mathcal{P}(\mathbb{C})$ sending a lattice Λ to the set of all quotients v_2/v_1, where (v_1, v_2) runs through all positively oriented bases of Λ. The positive orientation of the basis is equivalent to the condition $\Im v_2/v_1 > 0$, and therefore all these quotients lie the upper half-plane $\mathbb{H} = \{z \in \mathbb{C} : \Im z > 0\}$. The map ϕ is clearly blind to rotations

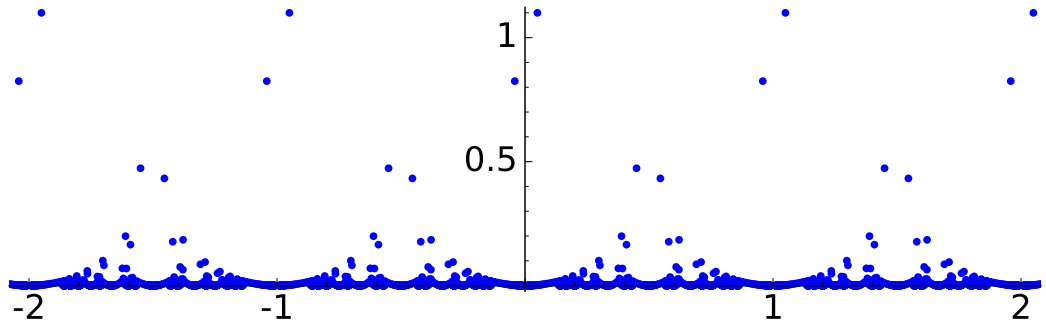


FIGURE 1.1. The set $\phi(\Lambda)$ where Λ is the lattice generated by $v_1 = (1, 0)$ and $v_2 = (0.05, 1.1)$.

and homotheties, inducing a quotient map (also denoted in the same way by abuse of notation) $\phi : \mathcal{L}/\mathbb{C}^* \to \mathcal{P}(\mathbb{C})$. This latter map turns out to be also injective, for if $c \in \phi(\Lambda_1) \cap \phi(\Lambda_2)$ we must have $c = u_2/u_1 = v_2/v_1$ for positively oriented bases (u_1, u_2) and (v_1, v_2) of Λ_1 and Λ_2, respectively, and therefore $\lambda = v_1/u_1$ satisfies $\lambda\Lambda_1 = \Lambda_2$. In figure 1.1 the reader can see an example of one of the sets $\phi(\Lambda)$.

To study the image of ϕ in more depth, fix a lattice Λ and consider two positively oriented bases (u_1, u_2) and (v_1, v_2). By the previous remarks we may find a matrix $\left(\begin{smallmatrix} d & c \\ b & a \end{smallmatrix}\right)$ in $\mathrm{SL}_2(\mathbb{Z})$ satisfying $v_1 = du_1 + cu_2$ and $v_2 = bu_1 + au_2$. Therefore

$$\frac{v_2}{v_1} = \frac{bu_1 + au_2}{du_1 + cu_2} = \frac{a\frac{u_2}{u_1} + b}{c\frac{u_2}{u_1} + d},$$

and the corresponding two points in the set $\phi(\Lambda)$ are related in this way. Motivated by this we define an action of the $\mathrm{SL}_2(\mathbb{Z})$ on the complex plane in the following way: the result of the action of a matrix $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ on a point z in the upper half-plane is given by

$$\gamma z = \frac{az + b}{cz + d}. \tag{1.1}$$

We have scrambled the rows and columns of the matrix to make it look pretty, but the idea is clear: any two points in $\phi(\Lambda)$ related by the action of a matrix γ correspond to bases (u_1, u_2) and (v_1, v_2) related by the matrix $\left(\begin{smallmatrix} d & c \\ b & a \end{smallmatrix}\right)$, which also lies in $\mathrm{SL}_2(\mathbb{Z})$, and viceversa. We conclude therefore that the image $\phi(\Lambda)$ is the orbit of a point in $\mathbb{H}$ under the action of $\mathrm{SL}_2(\mathbb{Z})$. Note also that every orbit is represented by some lattice, as $z \in \phi(\Lambda_z)$ where Λ_z is the lattice generated by 1 and $z \in \mathbb{H}$. In other words, ϕ induces a bijection $\mathcal{L}/\mathbb{C}^* \approx \mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$

That (1.1) defines an action follows from the interpretation we have just given, but it can also be checked directly. It is an easy computation to see that $\delta(\gamma z) = (\delta\gamma)z$ for matrices $\gamma, \delta \in \mathrm{SL}_2(\mathbb{Z})$. Another simple computation shows that

$$\Im\gamma z = \Im\frac{(az + b)(c\bar{z} + d)}{|cz + d|^2} = \frac{\Im z}{|cz + d|^2}. \tag{1.2}$$

In particular, $\Im z > 0$ if and only if $\Im\gamma z > 0$, and hence the action of $\mathrm{SL}_2(\mathbb{Z})$ leaves the upper half-plane $\mathbb{H}$ invariant.

This action is not faithful as the matrix $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ and its negative $-\gamma$ act exactly in the same way on $\mathbb{H}$. This is because both matrices lead to the same *fractional linear transformation* (or *Möbius transformation*) $z \mapsto (az + b)/(cz + d)$.

It is therefore more natural to consider the action as coming from the quotient group $\mathrm{SL}_2(\mathbb{Z})/\{\pm 1\}$. This group, called the *modular group*, is the one we want to study, although we might sometimes use this name also for $\mathrm{SL}_2(\mathbb{Z})$ by abuse of notation.

One of the motivations to consider the notion of lattices modulo $\mathbb{C}^*$ comes from the theory of quadratic forms. Recall that a primitive integral binary quadratic form of fundamental discriminant $d < 0$ which is positive definite can be obtained from a positively ordered basis of an ideal in the number ring of $\mathbb{Q}(\sqrt{d})$. These ideals are a very special kind of lattices in $\mathbb{C}$. Two of these lattices I and J generate properly equivalent quadratic forms if and only if $aI = bJ$ for some $a, b \in \mathcal{O}$ (which must have positive norm because $d < 0$). In this case, $I = \lambda J$ for $\lambda = a/b$, and reciprocally if $I = \lambda J$ for $\lambda \in \mathbb{C}$ then necessarily $\lambda \in \mathbb{Q}(\sqrt{d})$ and therefore $\lambda = a/b$ for some $a, b \in \mathcal{O}$. Hence the orbit $\phi(I)$ is an invariant of the proper equivalence class of quadratic forms derived from I: forms $Q(x, y) = |\alpha x + \beta y|^2/N(I)$ where (α, β) is a positively oriented basis of I. Note that if Q, as a function, is extended to $\mathbb{C}^2$ then $Q(x, y) = 0$ if and only if either $x = y = 0$, $x/y = -\beta/\alpha$ or $x/y = -\bar{\beta}/\bar{\alpha}$. We can use this to skip the ideals altogether, associating directly to the form Q the unique point $z_0 \in \mathbb{H}$ which is a zero of the polynomial $P(z)$ determined by $P(-x/y) = y^{-2}Q(x, y)$. If $Q(x, y) = ax^2+bxy+cy^2$, this point has the explicit expression $z_0 = (b+\sqrt{d})/(2a)$. Also by the previous remarks, two forms are properly equivalent if and only if their associated points lie in the same orbit modulo $\mathrm{SL}_2(\mathbb{Z})$. If we consider $\mathcal{F}_d$ the set of all the points $(b + \sqrt{d})/(2a)$ satisfying that a, b are integers, $a > 0$ and $4a \mid (b^2 - d)$, then each point in $\mathcal{F}_d$ determines a unique primitive form of discriminant d, by considering the minimal polynomial of this point appropriately scaled so that its coefficients are coprime integers. We have therefore the following correspondence

THEOREM. *Primitive quadratic forms of fundamental discriminant $d < 0$ which are positive definite are in one to one correspondence with points in $\mathcal{F}_d$. Properly equivalence classes are, from this point of view, orbits under the action of* $\mathrm{SL}_2(\mathbb{Z})$.

A similar theorem can be obtained relating indefinite primitive forms with a class of hyperbolic geodesics on the upper half-plane, but this is out of the scope of this exposition.[1]

A second motivation to study lattices modulo $\mathbb{C}^*$ comes from the theory of elliptic curves over $\mathbb{C}$. Any such curve can be constructed as a complex torus $\mathbb{C}/\Lambda$, where Λ is a lattice; and two elliptic curves $\mathbb{C}/\Lambda_1$ and $\mathbb{C}/\Lambda_2$ are isomorphic[2] if and only if the lattices Λ_1 and Λ_2 are related via multiplication by a nonzero complex constant. Due to the bijection induced by the map ϕ, the orbit space $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$ parametrizes all elliptic curves modulo isomorphism. Not only that, but there is also a natural notion of topology on this space —and even of Riemann surface— inherited from the one in $\mathbb{H}$. This makes the theory of elliptic curves much richer. Spaces like this one where each point represents an isomorphism class of some other object are ubiquitous in modern mathematics and receive the name of *moduli spaces* (modulus used as synonym of parameter).

[1]The interested reader can consult Siegel's article [**88**]. In this article Siegel proves an asymptotic formula previously stated by Gauss for the average of the class number for positive discriminants weightened by the logarithm of the fundamental unit. Although Siegel exploits the aforementioned relation between indefinite binary quadratic forms and hyperbolic geodesics, he was not the first one to discover it. Apparently this was first noted by Fricke and Klein in [**31**].

[2]Isomorphic means there is a bijective holomorphic map preserving the identity element for the group law. Such map is automatically a group homorphism.

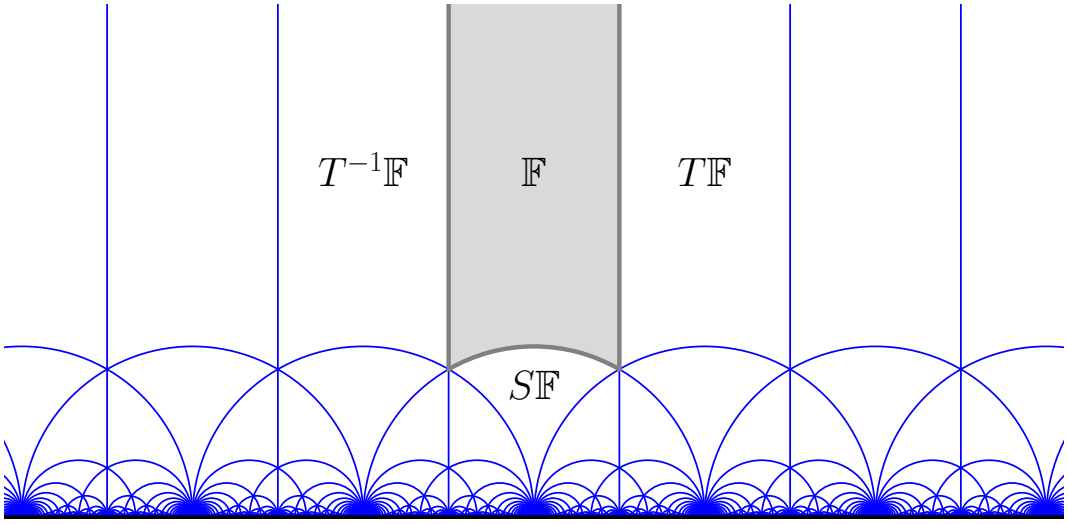


FIGURE 1.2. The fundamental domain $\mathbb{F}$ and its translations by the modular group: $\gamma(\mathbb{F})$ for $\gamma \in \mathrm{SL}_2(\mathbb{Z})/\{\pm 1\}$. The translations by the elements T, T^{-1} and S defined in (1.3) are labeled.

1.2. The fundamental domain

The action of the modular group $\mathrm{SL}_2(\mathbb{Z})/\{\pm 1\}$ on $\mathbb{H}$ is neither free nor transitive, but it is faithful. The action is also good in the sense that the orbits are discrete subsets of $\mathbb{H}$. When this happens, and the group acts by continuous transformations, it is often the case one can find a *fundamental domain*. This is a subset of the space a group is acting upon which contains exactly one point of every orbit (maybe with some exceptions) and which has nice topological properties, such as connectedness, etc. The definition is rather vague on purpose, and is often adapted to fit different contexts. In our case we will say that a region $\Omega \subset \mathbb{H}$ is a fundamental domain for the action of $\mathrm{SL}_2(\mathbb{Z})$ if it has finitely many connected components, each of them with piecewise smooth boundary, and satisfies that the translates $\{\gamma\Omega\}$ for $\gamma \in \mathrm{SL}_2(\mathbb{Z})/\{\pm 1\}$ cover the whole half-plane and only intersect on their boundaries. In other words, Ω *tiles* the half-plane by the action of the group $\mathrm{SL}_2(\mathbb{Z})$. If Ω satisfies the stronger property of containing exactly one point for every orbit, then we say Ω is a strict fundamental domain.[3] In any case, a fundamental domain is never unique, as for example all the translates by the group satisfy again the same properties. Which domain we choose to work with is up to us.

In our case we are going to choose the region

$$\mathbb{F} = \{z \in \mathbb{H} : |z| \geq 1,\ -1/2 \leq \Re z \leq 1/2\},$$

shown in figure 1.2, as the fundamental domain, as it is often done in the literature. We can make it strict by removing all the points in the boundary having negative real part; the resulting domain will be denoted by $\mathbb{F}'$. We are going to show $\mathbb{F}'$ is a strict fundamental domain in what follows.

Before we begin with the proof let us check that indeed the orbits are discrete. Suppose, by contradiction, that the orbit of some $z \in \mathbb{H}$ accumulates at some point $z_0 \in \mathbb{H}$, *i.e.* we can find a sequence of matrices $\gamma_n \in \mathrm{SL}_2(\mathbb{Z})$ such that $\lim_n \gamma_n z = z_0$ for some $z \in \mathbb{H}$ but $\gamma_n z \neq z_0$ for all n. Writing a_n, b_n, c_n and d_n for the entries of γ_n,

[3] In this case Ω and $\gamma\Omega$ may still intersect, but only at the fixed points of γ.

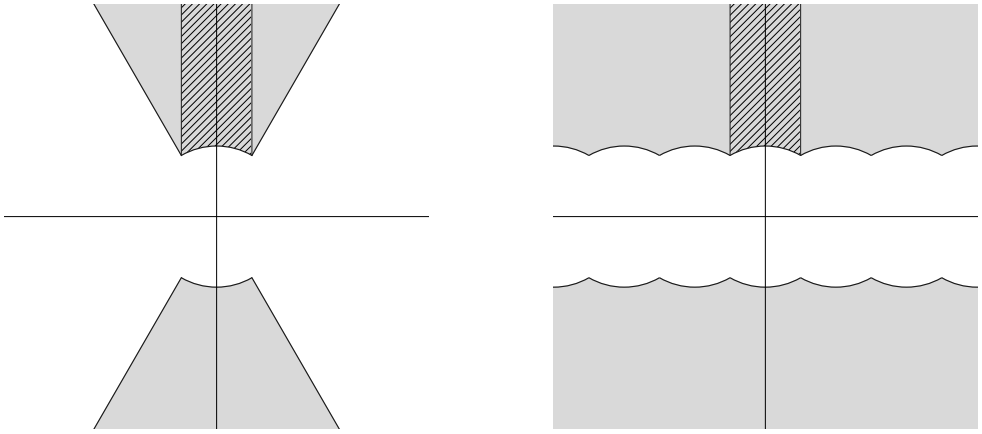


FIGURE 1.3. We show in grey the possible location of cz (left) and $cz+d$ (right) when $c \neq 0$ and $z \in \mathbb{F}$ (region with stripes). Note also that $|cz+d| = 1$ implies $|z| = 1$ and $|c| \leq 1$.

we establish by (1.2) the existence of the limit $\lim_n |c_n z + d_n| = (\Im z/\Im z_0)^{1/2} = \ell$. This also shows the existence of the limit $\lim_n |a_n z + b_n| = |z_0|\ell$. Now, $c_n z + d_n$ is a point lying in the lattice generated by 1 and z, and hence the modulus $|c_n z + d_n|$ may only take discrete values. For the limit to exist the value of $|c_n z + d_n|$ must stabilize for n big enough, and this only leaves finitely many possibilities for $c_n z + d_n$. Applying the same argument to $a_n z + b_n$, we conclude that $\gamma_n z$ may only take a finite number of values for n big enough, contradicting the fact that it accumulates at z_0.

Key to the proof are two very important elements of $\mathrm{SL}_2(\mathbb{Z})$, namely the matrices

$$S = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \qquad T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}. \tag{1.3}$$

The corresponding linear transformations, the inversion $Sz = -1/z$ and the translation $Tz = z + 1$, generate the modular group. Equivalently, $\mathrm{SL}_2(\mathbb{Z})$ is generated by $\{-1, S, T\}$. We will show this along the way. The proof is loosely based in the one given by Serre in [**85**].

We claim that by applying an appropriate combination of S and T we can move any point $z \in \mathbb{H}$ into the region $\mathbb{F}$. This can be done in the following way: we apply T or T^{-1} until $-1/2 \leq \Re z \leq 1/2$ and then, if not in $\mathbb{F}$, we apply S. Then we repeat as many times as necessary. Note that by (1.2) we have $\Im Sz = |z|^{-1}\Im z$ and therefore as long as $z \notin \mathbb{F}$ and $-1/2 \leq \Re z \leq 1/2$ we keep increasing $\Im z$. Either this process finishes or we obtain a sequence of points z_n satisfying $\Im z_n \to \alpha$ for some $\alpha > 0$ and $-1/2 \leq \Re z_n \leq 1/2$. In the latter case by compactness some subsequence must accumulate, contradicting the fact that the orbits are discrete. This establishes the claim. Therefore the process must finish. Note also that by applying T^{-1} (if $\Re z = 1/2$) or S (if $|z| = 1$ and $0 < \Re z \leq 1/2$) we can always move the point into the strict fundamental domain $\mathbb{F}'$.

Now assume two distinct points $z_1, z_2 \in \mathbb{F}$ are related by a matrix γ of $\mathrm{SL}_2(\mathbb{Z})$, *i.e.* $z_2 = \gamma z_1$. We claim that in this case these two points must have the same imaginary part, lie on the boundary of $\mathbb{F}$ and be symmetric with respect to the imaginary axis. To prove this, first note that rearranging the points if necessary we may assume $\Im z_2 \geq \Im z_1$. By (1.2) we have $\Im z_2 = \Im z_1/|cz_1 + d|^2$, but $|cz_1 + d| \geq 1$.

This is clear if $c = 0$, and otherwise cz_1 must lie in the region shown in the left part of figure 1.3 and therefore $cz_1 + d$ must lie in the one shown in the right. Hence $\Im z_2 = \Im z_1$ and $|cz_1 + d| = 1$. This latter fact, again by geometry, implies either $c = 0$ and $d = \pm 1$ or $c = \pm 1$ and $|z_1| = 1$. In the first case, necessarily $\gamma = \pm T^{\pm 1}$ and $|\Re z_1| = |\Re z_2| = 1/2$. In the second case, we may repeat the same analysis with $z_1 = \gamma^{-1} z_2$ to prove also $|z_2| = 1$. In both cases, if $z_1 \neq z_2$, the constraints force them to lie symmetrically with respect to the imaginary axis.

The two claims together show so far that $\mathbb{F}'$ contains one and only one point in each orbit, *i.e.* it is a strict fundamental domain. To show that the group $\mathrm{SL}_2(\mathbb{Z})/\{\pm 1\}$ is generated by S and T, we take any matrix $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ and consider the point $\gamma(2i)$, lying in $\mathbb{H}$. We have shown there is an element $\eta \in \mathrm{SL}_2(\mathbb{Z})$ in the subgroup generated by S and T satisfying $\eta\gamma(2i) \in \mathbb{F}'$, but then $\eta\gamma$ must fix $2i$. The coefficients of the matrix representing $\eta\gamma$ must then satisfy $a = d$, $b = -4c$ and $1 = a^2 + 4c^2$. This implies $\eta\gamma = \pm 1$ and therefore $\gamma = \pm\eta^{-1}$ can be written in terms of S, T and its inverses.

We can use the existence of the fundamental domain to give a short proof of the finiteness of the class number for a fundamental discriminant $d < 0$. Note that by the previous remarks on the topic it suffices to show that $\mathcal{F}_d \cap \mathbb{F}$ contains finitely many points. This is a simple verification: any element of $\mathcal{F}_d$ is of the form $(b + \sqrt{d})/(2a)$, and the restriction on the imaginary part $\sqrt{|d|}/(2a) \geq \sqrt{3}/2$ limits the possible values of a, while for each of these the restriction on the real part $-1/2 \leq b/(2a) \leq 1/2$ limits the possible values of b. If one translates more carefully the condition of the point lying in $\mathbb{F}'$ to the coefficients a, b and $c = (b^2 - d)/(4a)$ we recover the following theorem by Gauss:

THEOREM (GAUSS). *Every positive definite primitive quadratic form is properly equivalent to one and only one form $ax^2 + bxy + cy^2$ satisfying either*

$$-a < b \leq a < c \qquad or \qquad 0 \leq b \leq a = c.$$

The procedure described above to move any point of $\mathbb{H}$ into $\mathbb{F}'$ also tells us an algorithmic way to compute this standard representative.

1.3. Continued fractions and the group structure

Continued fractions provide a system to represent real numbers different from the usual decimal (or n-ary) expansions. The number is represented in the form $[a_0; a_1, \ldots, a_n, \ldots]$, where a_0 is an arbitrary integer and the rest of the a_i are strictly positive integers, but not necessarily bounded. The main advantage is that a lot of Diophantine approximation properties may be read from these coefficients.

Informally, the continued fraction expansion $[a_0; a_1, \ldots, a_n, \ldots]$ represents the real number

$$a_0 + \cfrac{1}{a_1 + \frac{1}{a_2 + \cdots}}.$$

One way to formalize this is to define inductively $[x] = x$, $[a_0; x] = a_0 + 1/x$, $[a_0; a_1, x] = a_0 + 1/(a_1 + 1/x)$ and, in general,

$$[a_0; a_1, \ldots, a_n, x] = [a_0; a_1, \ldots, a_{n-1}, a_n + 1/x].$$

We also define $[a_0; a_1, a_2, \ldots] = \lim_n [a_0; a_1, \ldots, a_n]$, limit which always exists and which may attain any real value depending on the coefficients. The proof of these facts can be consulted in most number theory treatises, for example [**68**] or chapter X

of [**46**]. We are going to include here some simple proofs concerning finite expansions which will be used to obtain a presentation for the modular group. More concretely, we are going to prove that any rational number admits a unique expression as a finite continued fraction $[a_0; a_1, \dots, a_n]$ where a_0 is an arbitrary integer, the rest of the a_i are strictly positive integers and either $n = 0$ or $a_n \geq 2$.

Let us show first the uniqueness. By induction, we have for $n \geq 1$,

$$[a_0; a_1, \dots, a_n] = [a_0; y] \qquad \text{where} \qquad y = [a_1; a_2, \dots, a_n]. \tag{1.4}$$

Using this and induction again we obtain the inequalities

$$a_0 \leq [a_0; a_1, \dots, a_n] < a_0 + 1, \tag{1.5}$$

where the first inequality is strict for $n \geq 1$. Suppose now $[a_0; a_1, \dots, a_n] = [b_0; b_1, \dots, b_m]$ where $n \leq m$. By (1.5) we must have $a_0 = b_0$, and if $n = 0$ necessarily $m = 0$ and we are finished. Otherwise, by (1.4), the tails coincide $[a_1; a_2, \dots, a_n] = [b_1; b_2, \dots, b_m]$ and an inductive argument finishes the proof.

We give now an algorithm closely related to Euclid's showing the existence of such expansion. Since $[a_0; a_1, \dots, a_n] - k = [a_0 - k; a_1, \dots, a_n]$ we will only be concerned with positive rational numbers. Let $x = p/q$ where p and q are positive coprime integers. The algorithm will consist in applying the map $x \mapsto x - 1$ until $p < q$ (transforming p/q to its fractional part $\{p/q\}$), then performing the inversion $x \mapsto 1/x$ and repeating. Note the quantity $p + q$ keeps decreasing and therefore we are guaranteed to arrive sooner or later to $x = 0$. Once this happens we may apply the operations in reverse order to $x = 0$ to obtain a continued fraction expression for our original rational. To see this note that the two maps involved are provided by the linear fractional transformations associated to T, defined in (1.3), and to

$$\bar{S} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

closely related to S also defined in (1.3). The algorithm therefore provides a sequence of non-negative integers $a_0, \dots, a_n$, where only a_0 may vanish, satisfying

$$T^{-a_n}\, \bar{S}\, T^{-a_{n-1}}\, \bar{S} \cdots \bar{S}\, T^{-a_1}\, \bar{S}\, T^{-a_0} x = 0.$$

Using $\bar{S}^{-1} = \bar{S}$, this is equivalent to

$$x = T^{a_0}\, \bar{S}\, T^{a_1}\, \bar{S} \cdots \bar{S}\, T^{a_n} 0, \tag{1.6}$$

and by (1.4) and induction this is also equivalent to $x = [a_0; a_1, \dots, a_n]$. Finally we need to "fix" the expansion if $a_n = 1$ and $n \neq 0$. In this case we use the identity

$$[a_0; a_1, \dots, a_{n-1}, 1] = [a_0; a_1, \dots, a_{n-1} + 1].$$

The matrix $\bar{S}$ we introduced does not lie in $\mathrm{SL}_2(\mathbb{Z})$ as it has determinant -1, but with the help of the identities

$$\bar{S}\, T^n\, \bar{S} x = S\, T^{-n}\, S x \qquad \text{and} \qquad \bar{S}\, T^n 0 = S\, T^{-n} 0$$

we can rewrite (1.6) as

$$x = T^{a_0}\, S\, T^{-a_1}\, S\, T^{a_2}\, S \cdots S\, T^{(-1)^n a_n} 0. \tag{1.7}$$

Using this and the uniqueness result we are going to show that the modular group can be presented as

$$\mathrm{SL}_2(\mathbb{Z})/\{\pm 1\} = \langle\, S, T \mid S^2 = (ST)^3 = 1 \,\rangle.$$

For this it suffices to show that given any word w in S and T whose associated linear fractional transformation evaluates to the identity, we can use the two given relations to transform the word itself to the identity. By using $S^2 = 1$ and grouping the T elements together we can always assume that w has the form

$$w = T^{b_0} S T^{b_1} S \cdots S T^{b_n} \tag{1.8}$$

for some integers $b_i \in \mathbb{Z}$, where only b_0 and b_n may vanish. First we show that we can use the given relations to transform the word in such a way as to guarantee that the sign of b_i coincides with $(-1)^i$ for $1 \leq i \leq n$, or for $1 \leq i \leq n-1$ if $b_n = 0$. We use induction on n. If $n = 0$ the condition is void, and if $n = 1$ the only non-trivial case has $b_1 \geq 1$. If so we may write

$$w = T^{b_0-1} T S T T^{b_1-1} = T^{b_0-1} S T^{-1} S T^{b_1-1},$$

where we have used the identity $T S T = S T^{-1} S$. The latter word satisfies the hypothesis.

Assume now the result is true for $n-1$. Applying the induction hypothesis to w (and updating the value of n if necessary) we may assume that only b_n in (1.8) has the wrong sign. If n is odd this means that $b_n \geq 1$. Write

$$\begin{aligned} w &= T^{b_0} S \cdots S T^{b_{n-1}-1} T S T T^{b_n-1} \\ &= T^{b_0} S \cdots S T^{b_{n-1}-1} S T^{-1} S T^{b_n-1}, \end{aligned}$$

where we have used again $T S T = S T^{-1} S$. Since b_{n-1} had the appropriate sign, $b_{n-1} \geq 1$. If $b_{n-1} \geq 2$ this word satisfies the requeriments. Otherwise $b_{n-1} = 1$ and

$$w = T^{b_0} S \cdots S T^{b_{n-2}-1} S T^{b_n-1}.$$

Again b_{n-2} had the appropriate sign and therefore $b_{n-2} \leq -1$. Hence this word satisfies the requeriments. Finally, if n was even instead of odd, the same argument using $T^{-1} S T^{-1} = S T S$ instead works, taking special care if $n = 2$. Once this has been established we may rename $a_i = (-1)^i b_i$, $a_n = (-1)^n (b_n + m)$, multiply on the right by T^m and evaluate w at 0 to obtain the identity

$$T^{a_0} S T^{-a_1} S \cdots S T^{(-1)^n a_n} 0 = T^m 0.$$

We may choose m so that these two continued fraction expansions are under the above hypothesis for the uniqueness result to hold, and therefore we must have $n = 0$ and $a_0 = m$, effectively showing that $w = 1$.

Throughout this proof we have been implicitely extending the action of $\mathrm{SL}_2(\mathbb{Z})$ to certain rational numbers. It is possible to extend it to all of them, but we have to take into account the possibility of evaluating a linear fractional transformation at its pole. The trick is to make the group act on the set $\mathbb{Q} \cup \{\infty\}$ in the natural way: given $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ and $x \in \mathbb{Q}$ then γx is the result of evaluating the linear fractional transformation at x, except when x is its pole, case in which we define $\gamma x = \infty$. We also define $\gamma\infty = \lim_{x\to\infty} \gamma x$; this is, $\gamma\infty = a/c$ if $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ and $c \neq 0$ and $\gamma\infty = \infty$ if $c = 0$. Using Bézout's identity it is immediate that the action thus defined on $\mathbb{Q} \cup \{\infty\}$ is transitive. The stabilizer of any point is therefore conjugated to the stabilizer of ∞, which coincides with the subgroup generated by T. Another useful fact is that if $\gamma x \neq \infty$ and $x = p/q$ for coprime p and q then $\gamma x = (ap+bq)/(cp+dq)$, where $ap + bq$ and $cp + dq$ are, again, coprime.

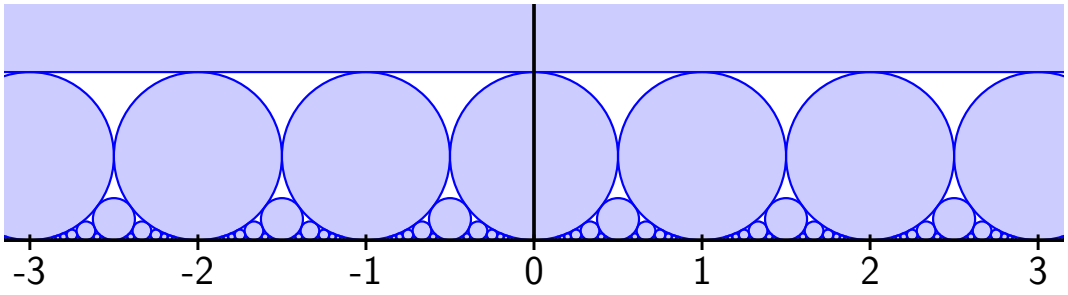


FIGURE 1.4. The Ford circles.

1.4. Ford circles

The actions we have defined of $\mathrm{SL}_2(\mathbb{Z})$ on $\mathbb{H}$ and on $\mathbb{Q}\cup\{\infty\}$ are, of course, related. One way to make this explicit is by the use of *Ford circles*, introduced by Ford in the beautifully written article [**30**]. These are defined as follows: for every rational p/q, where p and q are coprime integers, $q\geq 1$, we associate the circle of radius $1/(2q^2)$ tangent to the real line at the rational, *i.e.* centered at $p/q+i/(2q^2)$. We also associate a degenerate "circle" to ∞ consisting of all points $z\in\mathbb{H}$ with $\Im z\geq 1$. These circles, shown in figure 1.4, turn out to be either disjoint or tangent to each other (see [**30**] for a proof), and they also have the important property of being preserved by the action of $\mathrm{SL}_2(\mathbb{Z})$. We are going to need in fact a slightly more general result: if we denote by $\mathcal{F}_{p/q}(\delta)$ the circle of radius $\delta/(2q^2)$ tangent to the real line at p/q, and by $\mathcal{F}_\infty(\delta)$ the region $\{\Im z\geq\delta^{-1}\}$, we also have

$$\gamma(\mathcal{F}_x(\delta))=\mathcal{F}_{\gamma x}(\delta) \tag{1.9}$$

whenever $\gamma\in\mathrm{SL}_2(\mathbb{Z})$ and $x\in\mathbb{Q}\cup\{\infty\}$. The sets $\mathcal{F}_x(\delta)$ receive the name of *generalized Ford circles* or *Speiser circles*.

LEMMA 1.1. *Let p, q be coprime integers and $\delta>0$. Given $z\in\mathbb{H}$, the following conditions are equivalent:*

(*i*) $z\in\mathcal{F}_{p/q}(\delta)$.

(*ii*) $|qz-p|^2\leq\delta\Im z$.

(*iii*) $\gamma z\in\mathcal{F}_\infty(\delta)$ *for any* $\gamma\in\mathrm{SL}_2(\mathbb{Z})$ *satisfying* $\gamma(p/q)=\infty$.

It is to be understood that $p/q=\infty$ if $q=0$.

PROOF. If $q=0$ then $p=\pm1$ and $\gamma=\pm T^n$, and the equivalences are trivial. If $q\neq0$, writing $z=x+iy$ and squaring, (*i*) is equivalent to

$$\left(x-\frac{p}{q}\right)^2+\left(y-\frac{\delta}{2q^2}\right)^2\leq\frac{\delta^2}{4q^4}.$$

Expanding the second square and multiplying by q^2 it is clear that (*i*) $\iff$ (*ii*). The equivalence (*ii*) $\iff$ (*iii*) follows from formula (1.2) after noting that $\gamma=\pm\left(\begin{smallmatrix}*&*\\-q&p\end{smallmatrix}\right)$. □

We proceed to prove (1.9) now by cases. If either $x=\infty$ or $\gamma x=\infty$, then the identity follows from (*i*) $\iff$ (*iii*) applying the lemma either to γ or γ^{-1}. If $x\neq\infty\neq\gamma x$, we can choose $\eta\in\mathrm{SL}_2(\mathbb{Z})$ satisfying $\eta x=\infty$ and apply the previous case to show

$$\gamma(\mathcal{F}_x(\delta))=(\gamma\eta^{-1})(\mathcal{F}_\infty(\delta))=\mathcal{F}_{\gamma x}(\delta).$$

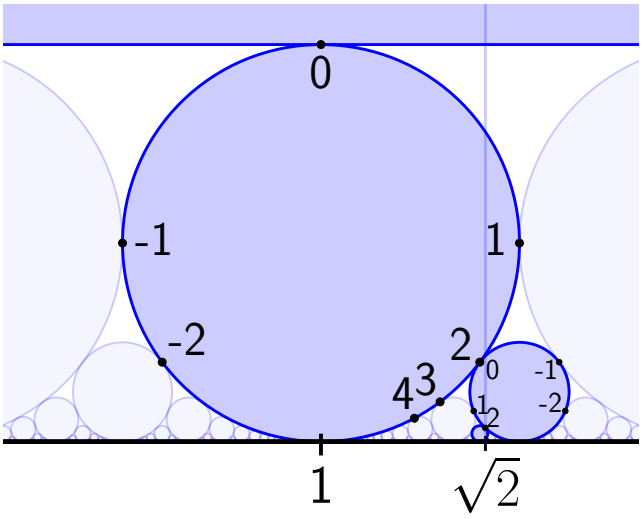


FIGURE 1.5. The Ford circles intersecting the vertical line over $\sqrt{2} = [1; 2, 2, \cdots]$ appear highlighted. These intersected Ford circles lie over the consecutive convergents: $1 = [1]$, $3/2 = [1; 2]$, $7/5 = [1; 2, 2]$, ...

COROLLARY 1.2. *For $\delta = 1$ the Ford circles $\mathcal{F}_x(\delta)$ are either disjoint or tangent. For $\delta \geq 2$ the Speiser circles $\mathcal{F}_x(\delta)$ cover the upper half-plane.*

PROOF. Some elementary geometry shows that the distance between the centers of $\mathcal{F}_{p/q}(1)$ and $\mathcal{F}_{P/Q}(1)$ is given by

$$\left(\frac{1}{2q^2} + \frac{1}{2Q^2}\right) + \frac{(Pq - pQ)^2 - 1}{Q^2 q^2}. \tag{1.10}$$

If $p/q \neq P/Q$ then $|Pq - pQ| \geq 1$ showing that the circles are either tangent or disjoint.

For $\delta \geq 2$ note that the fundamental domain $\mathbb{F}$ is contained in $\mathcal{F}_\infty(\delta)$. Since $\mathbb{F}$ covers the plane when translated by the modular group, so does $\mathcal{F}_\infty(\delta)$. □

As we vary δ the identity (1.9) gives us a fairly good sense of how $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ acts on the upper half-plane once we know how it acts on $\mathbb{Q} \cup \{\infty\}$. In particular, the change of variables $w = \gamma z$ moves the region $\{\Im z \geq \delta^{-1}\}$ to the the Ford circle $\mathcal{F}_{\gamma\infty}(\delta)$ in the w-variable, and hence if $f : \mathbb{H} \to \mathbb{C}$ is any complex-valued function, the function $g(z) = f(\gamma z) = f(w)$ behaves when $\Im z \to \infty$ as f does as $w \to \gamma\infty$ within the generalized Ford circles. If we choose any other $\gamma_1 \in \mathrm{SL}_2(\mathbb{Z})$ satisfying also $\gamma_1\infty = \gamma\infty$, then $\gamma_1 = \pm\gamma T^n$ for some $n \in \mathbb{Z}$ and hence $g_1(z) = f(\gamma_1 z) = g(z + n)$ is just a translation of g. If $\gamma_1\infty = x \in \mathbb{Q}$ then note that $\gamma_1 S0 = x$ and therefore $\gamma_1 S$ admits a decomposition in S and T of the form (1.7) whose coefficients for some n are precisely the coefficients prescribed by the continued fraction expansion of x. Therefore, in some sense, moving the ford circle over x to the Ford circle at infinity by the action of the modular group requires applying translations and inversions in a precise order to "undo" the continued fraction expansion.

It is due to Ford [**30**] that this process can also be read from the circles directly (for $\delta = 1$), with no intervention of the modular group. To see this, note that the element $\gamma = T^n S$ moves $\mathcal{F}_\infty(1)$ to $\mathcal{F}_n(1)$, *i.e.* changing the value of n determines in which of the tangent circles to $\mathcal{F}_\infty(1)$ we end up (see figure 1.4). Now, since it preserves the Ford circles, the application $w = \gamma z$ must also send every tangent

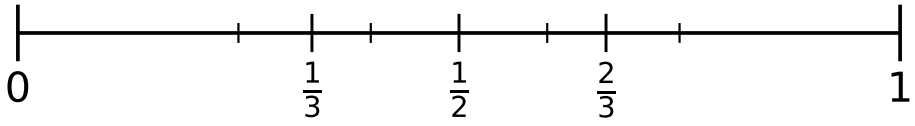


FIGURE 1.6. The Farey dissection of order 3. The members of the Farey sequence are labeled, while their medians are indicated by ticks.

circle to $\mathcal{F}_\infty(1)$ to a tangent circle to $\mathcal{F}_n(1)$, so if z already lies in one of the former, γz will lie in one of the latter. Hence if $z = T^m S u$ with $u \in \mathcal{F}_\infty(1)$ then choosing m we can determine in which of the circles tangent to $\mathcal{F}_n(1)$ the variable w ends up. We can write $w = T^n S T^m S u$ and iterate this process: $u = T^k S v$ to reach all the circles which are tangent to a tangent circle to $\mathcal{F}_n(1)$, etc. From (1.7) we see that the continued fraction coefficients of the rational x are coordinates specifying the path to take if we want to go from $\mathcal{F}_\infty(1)$ to $\mathcal{F}_x(1)$ jumping from tangent Ford circle to tangent Ford circle. The coordinate system is laid as follows: 0 refers to the tangent circle we came from, and then $1, 2, \ldots$ specify consecutive tangent circles in one direction and $-1, -2, \ldots$ in the other direction. If the circles are counted clockwise or counter-clockwise depends on the parity of the number of circles we have already traveled. The circles visited are also characterized by being those who intersect the ray orthogonal to the real line at x, hence the continued fraction coefficients also give the "tangency coordinates" specifying the circles we encounter as we descend from $x + i\infty$ through this ray. This interpretation is also valid for any real number, and for example we can see in figure 1.5 how to interpret the first coefficients for the irrational $\sqrt{2}$.

The rationals associated to the Ford circles we visit when descending toward a real number correspond to the partial continued fractions $[a_0]$, $[a_0; a_1]$, etc. These are called the *convergents*, and always provide rationals which best approximate the real number among all rationals with the same of smaller denominator; this is clear from the geometry of the Ford circles.[4] Note that when x is irrational we intersect infinitely many circles, recovering the classical Dirichlet's theorem that states that there infinitely-many rationals p/q satisfying $|x - p/q| \leq q^{-2}$. Speiser circles lead to some refinements of this theorem, as the δ parameter is related to how close all the points in the circle are to the base rational. In particular they can be used to give a short proof of Hurwitz's theorem, which states that if we replace the inequality $|x - p/q| \leq q^{-2}$ with the stronger statement $|x - p/q| \leq Cq^{-2}$ then the same result holds for every irrational number x if and only if $C \geq 1/\sqrt{5}$ (see [**30**]).

1.5. The Farey sequence

Given any integer N, the *Farey sequence of order* N refers to all the rational numbers in $[0, 1]$ having denominator bounded above by N, arranged in order of increasing size. These rational points can also be characterized as those lying at the base of all the Ford circles $\mathcal{F}_{p/q}(1)$ intersected by the segment $L_N = \{0 \leq x \leq 1, y = 1/N^2\}$. They satisfy the following remarkable properties:

PROPOSITION 1.3. *Let $p/q < P/Q$ be two consecutive rationals in the Farey sequence of order N, written in their lowest terms. Then*

[4]For other denominators best-approximants are always mediants of convergents (theorem 15 of [**68**]).

(*i*) $Pq - pQ = 1$.

(*ii*) $N + 1 \leq q + Q \leq 2N$.

(*iii*) $$\frac{p+P}{q+Q} - \frac{p}{q} = \frac{1}{q(q+Q)} \qquad \text{and} \qquad \frac{P}{Q} - \frac{p+P}{q+Q} = \frac{1}{Q(q+Q)}.$$

PROOF. If any curve intersects two Ford circles in succession these must be tangent, as the only way to leave a circle is through a tangency point, or through the common boundary with a "curved triangle" whose two other sides correspond to tangent circles (see figure 1.4, for a formal proof this fact it can be shown to be true when leaving $\mathcal{F}_\infty(1)$ and then it must hold when leaving any other circle by transforming the half-plane under the action of the modular group). Therefore any two Ford circles intersected in succession by the segment L_N are tangent, and formula (1.10) shows (*i*) must hold. The rational $(p+P)/(q+Q)$ lies strictly in between p/q and P/Q and therefore is not part of the Farey sequence. Hence $q + Q \geq N + 1$, and trivially $q + Q \leq 2N$. Finally (*iii*) follows from (*i*). □

It is often the case we want to dissect the segment L_N into smaller intervals, in a way that every of the subintervals is appropriately close a rational p/q in the sense that it is contained in $\mathcal{F}_{p/q}(\delta)$ for some fixed δ. If $\delta < 2/\sqrt{3}$ then this is impossible because the Ford circles do not cover the half-plane, but if $\delta > 1$ then the intersections of L_N with the Speiser circles might overlap. A "clean" way to do this is the *Farey dissection of order N*. We associate to each rational p/q in the Farey sequence of order N the interval

$$\mathcal{A}_{p/q} = \left[\frac{p+p^-}{q+q^-}, \frac{p+p^+}{q+q^+}\right),$$

where $p^-/q^- < p/q < p^+/q^+$ are consecutive rationals in this sequence (figure 1.6). What we do with the endpoints of the interval is a matter of convenience, in our case it will be useful to consider two half-intervals $\mathcal{A}_0 = [0, 1/(N+1))$ and $\mathcal{A}_1 = [N/(N+1), 1]$. The intervals $\mathcal{A}_{p/q}$ are disjoint and cover $[0, 1]$. Moreover

$$\mathcal{F}_{p/q}(1/4) \cap L_N \subset \mathcal{A}_{p/q} + i/N^2 \subset \mathcal{F}_{p/q}(2) \cap L_N. \tag{1.11}$$

To see this, let x be one of the edge points of $\mathcal{A}_{p/q}$ and $z = x + i/N^2$. Then a simple computation shows $|qz - p|^2 = \delta(\Im z)$ for $\delta = N^2/(q+Q)^2 + q^2/N^4$ which by proposition 1.3 lies in $[1/4, 2]$. By (*ii*) of lemma 1.1 above this is equivalent to the inclusions (1.11).

The concept of the Farey dissection trivially generalizes to other intervals and to the continuum, considering the intervals associated to rationals p/q with $q \leq N$ in the given interval or in the whole real line.

1.6. Geometry

We cannot finish this section without saying a few words about Poincaré's model of hyperbolic geometry in the upper half-plane. If we endow $\mathbb{H}$ with the arclength element

$$ds = \frac{\sqrt{(dx)^2 + (dy)^2}}{y} \qquad \text{where} \quad z = x + iy,$$

we obtain a riemannian manifold of constant curvature -1, where the group of orientation-preserving isometries can be identified with $\mathrm{SL}_2(\mathbb{R})/\{\pm 1\}$ with the usual

action.[5] In particular all elements of the modular group preserve all the features of the geometry on $\mathbb{H}$, some of which described in what follows. Geodesics are either vertical rays or half-circles whose center lies on the real line, while the angles coincide with the Euclidean ones. In this sense the fundamental domain $\mathbb{F}$ is a hyperbolic triangle, with inner angles $\pi/3$, $\pi/3$ and 0 and a missing vertex. The missing vertex can be identified with the point ∞ in the "boundary". In fact, as a topological space the whole model can be compactified by adding the set of end-points of all geodesics or *limit points* $\mathbb{R} \cup \{\infty\}$. This construction is analogous to that of the projective plane for the Euclidean plane. All the linear fractional transformations have a well-defined action in this new space.

When seen in the Riemann sphere via the stereographic projection the upper half-plane then becomes a disk on the sphere, and the set of limit points its boundary. In fact, one can define an appropriate arclength element on the usual open unit disk so that the resulting geometry is equivalent to the one described above and the set of limit points coincides with the boundary of the disk. This is Poincaré's disk model. The metric spaces obtained when leaving the limit points aside are isometric and, in fact, conformally equivalent in the usual sense, as in both cases the notion of angle coincides with the Euclidean one. We will directly work with the disk model, but it is important to keep in mind that the set of limit points $\mathbb{R} \cup \{\infty\}$ is topologically a circle and in this sense the linear fractional transformations act continuously on it.

The orientation-preserving isometries on the upper half-plane can be classified depending on the number and location of their fixed points (in a similar way to what happens in the Euclidean plane). To describe this, let $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{R})$ be distinct from ± 1. The fractional linear transformation associated to γ fixes the point ∞ if and only if $c = 0$, and in this case the other fixed point is given by $-b/(a-d)$. If, on the contrary, $c \neq 0$ then the fixed points are given by the expression

$$\frac{a - d \pm \sqrt{\Delta}}{2c} \qquad \text{where} \quad \Delta = (a+d)^2 - 4. \tag{1.12}$$

Note that in any case the sign of Δ determines the nature of the fixed points: if $\Delta < 0$ then γ has one fixed point lying in $\mathbb{H}$, if $\Delta = 0$ then γ has one fixed point lying in $\mathbb{R} \cup \{\infty\}$ and if $\Delta > 0$, γ has two fixed points lying in $\mathbb{R} \cup \{\infty\}$. In the first case we say the transformation is *elliptic*, in the second case *parabolic* and in the third *hyperbolic*. There are isometries fixing any of these combinations of points, and once the fixed points are chosen the resulting isometries form an uniparametric group isomorphic to S^1 in the elliptic case and to $\mathbb{R}$ in the other two. In figure 1.7 some examples of the orbits by these uniparametric groups can be seen.

Forcing the coefficients of the matrix to be integers imposes strong conditions on the nature of the fixed points, as the following theorem shows.

THEOREM 1.4. *If a transformation in* $\mathrm{SL}_2(\mathbb{Z})$ *is parabolic, the fixed point lies in* $\mathbb{Q} \cup \{\infty\}$. *If it is hyperbolic the fixed points are always a pair of Galois-conjugated quadratic surds. If it is elliptic the fixed point is always in the orbit of* i *or* $\rho = (1 + i\sqrt{3})/2$ *modulo* $\mathrm{SL}_2(\mathbb{Z})$. *There are transformations in* $\mathrm{SL}_2(\mathbb{Z})$ *fixing any of the above specified points.*

[5] If we add the map $z \mapsto -\bar{z}$ we obtain the whole group of isometries. This group can be seen to be isomorphic to $\mathrm{S^*L}_2(\mathbb{R})/\{\pm 1\}$, where $\mathrm{S^*L}_2(\mathbb{R})$ stands for the group of 2×2 matrices with real entries and determinant ± 1. Under this isomorphism, the elements of negative determinant act via the corresponding linear fractional transformation composed with complex conjugation.

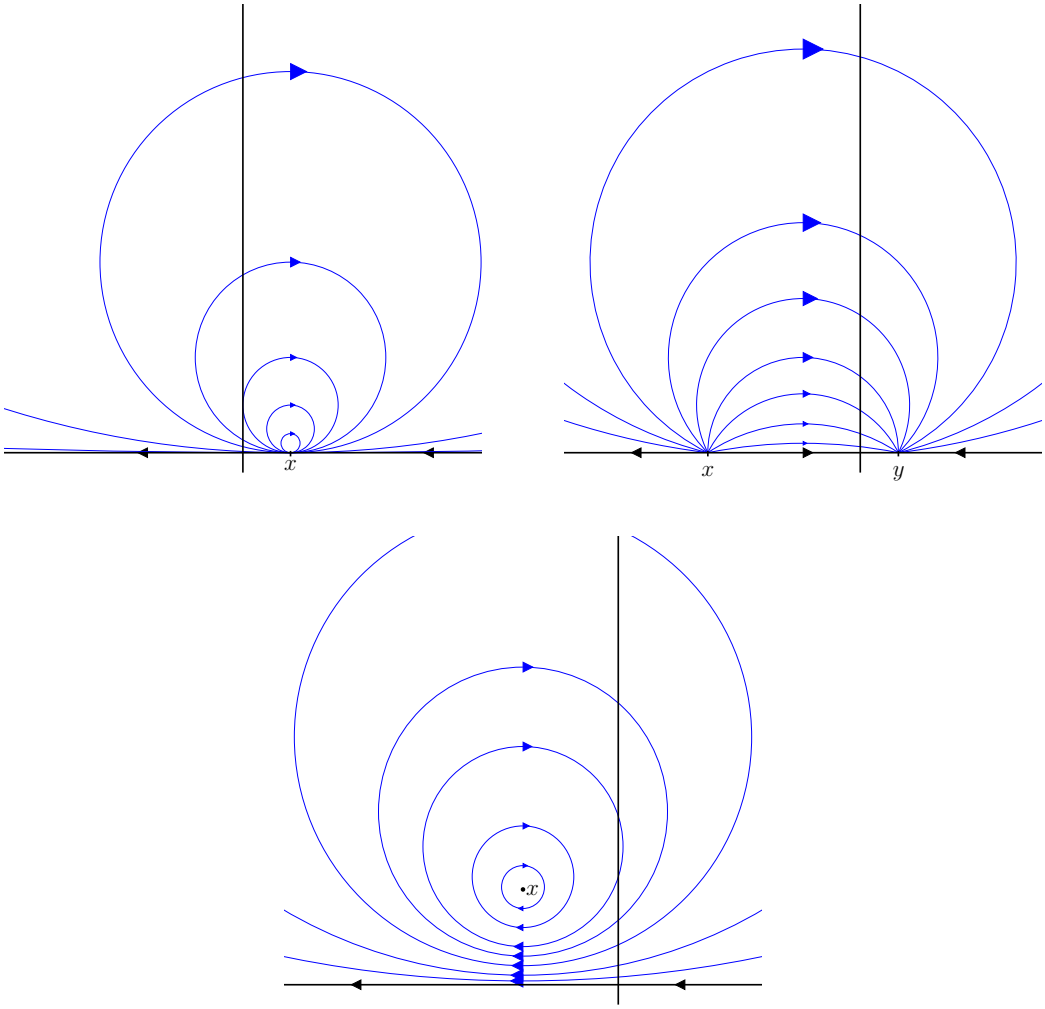


FIGURE 1.7. Examples of orbits when Poincaré's upper half-plane is acted by uniparametric groups of orientation-preserving isometries. In these examples the orbits are always circles. Two particular cases are missing: parabolic transformations fixing ∞ are horizontal translations, while hyperbolic transformations fixing ∞ and x are Euclidean dilations fixing the point x.

PROOF. We recall the only transformations fixing ∞ are the translations, which are parabolic. Since the action of $\mathrm{SL}_2(\mathbb{Z})$ on $\mathbb{Q} \cup \{\infty\}$ is transitive, there are parabolic transformations fixing any rational point, and any transformation not equal to the identity fixing a rational point must also be parabolic.

We can therefore assume $c \neq 0$, $\Delta \neq 0$ and that the fixed points are given by (1.12). Note that Δ, if positive, cannot be a square as it would contradict the previous remark. Therefore when the transformation is hyperbolic the pair of fixed points are complex-conjugated quadratic surds. To see that any pair of quadratic surds are fixed by some hyperbolic transformation we appeal to the fact that any quadratic surd has a periodic expansion as a continued fraction (theorem 177 of [**46**]). This means, in view of (1.7) that there exist $\gamma, \eta \in \mathrm{SL}_2(\mathbb{Z})$ satisfying that $\lim_n \eta\gamma^n 0 = x$, where x is one of the quadratic surds. Hence $x = \lim_n \eta\gamma^{n+1}0 = \eta\gamma\eta^{-1}x$ or $\eta\gamma\eta^{-1}$ fixes x, and $\eta\gamma\eta^{-1} \neq \pm 1$ as otherwise $x = \eta 0$ would be rational.

Assume finally we are in the elliptic case. The transformation S fixes i, while TS fixes ρ, and hence all points in these two orbits are admissible. If, on the other hand, we are given a transformation which fixes a point in $\mathbb{H}$, we can conjugate it by an element of $\mathrm{SL}_2(\mathbb{Z})$ to assume without loss of generality that the fixed point lies in $\mathbb{F}'$. The same proof we have given showing that $\mathbb{F}'$ is a strict fundamental domain now shows that the fixed point has modulus 1, and that the transformation satisfies $c = \pm 1$. By (1.12) this implies that the real part of the fixed point $\pm(a-d)/2$ must be an integer multiple of $1/2$, hence only leaving i and ρ as possibilities. □

Given any subgroup of $\mathrm{SL}_2(\mathbb{R})$, the points fixed by parabolic transformations are called *cusps*, while those fixed by elliptic transformations are called *elliptic points*. This theorem shows that for $\mathrm{SL}_2(\mathbb{Z})$ the cusps are $\mathbb{Q} \cup \{\infty\}$, while the elliptic points are the orbits of i and ρ. For any finite index subgroup of $\mathrm{SL}_2(\mathbb{Z})$ the cusps are again the same set, as for any parabolic transformation lying in $\mathrm{SL}_2(\mathbb{Z})$ some power lies in the subgroup, fixes the same point, and cannot be the identity as these transformations are not of finite order. The set of points fixed by hyperbolic transformations is also preserved by the same argument, but some elliptic points may dissapear.

CHAPTER 2

Classical modular forms

In this chapter we introduce the concept of classical modular forms for arbitrary finite index subgroups of the modular group and arbitrary multiplier systems. We have to be quite selective regarding the results we present, as this topic is vast and admits many generalizations in many different directions. We are going to start by introducing the simplest case: modular forms for the whole modular group and trivial multiplier system —although this was not historically the first case studied, as Jacobi's theta function is not of this kind— and then incrementally complicate the definition.

A good basic reference for the simplest case is Serre's book [**85**], while the first chapter of [**97**] by Zagier provides a good survey on the different existing generalizations. Treatises covering in depth the analytic aspects are provided by Rankin [**82**] and Iwaniec [**61**].

2.1. Classical modular forms for $\mathrm{SL}_2(\mathbb{Z})$

Let $k \in \mathbb{R}$. A *modular function of weight k* is an analytic function $f : \mathbb{H} \to \mathbb{C}$ satisfying the following invariance relation under the action of the modular group:

$$f(\gamma z) = (cz + d)^k f(z) \qquad \text{for every} \quad \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}). \tag{2.1}$$

Note that since γ and $-\gamma$ act in the same way, if f is not identically zero this immediately implies that k is an even integer. The value of k is called the *weight* of the modular function.

Of course since the modular group is generated by S and T, equation (2.1) is equivalent to the pair of conditions

$$\begin{cases} f(z+1) = f(z), \\ f(-1/z) = z^k f(z). \end{cases} \tag{2.2}$$

The first one, in particular, shows that if we perform the change of variables $q = e^{2\pi i z}$ the function $g(q) = f(z)$ is well-defined and analytic on the punctured unit disk. It therefore admits a Laurent expansion, which can be translated back to a Fourier expansion for f:

$$f(z) = \sum_{n=-\infty}^{\infty} a_n q^n = \sum_{n=-\infty}^{\infty} a_n e^{2\pi i n z}.$$

We say f is a *modular form* if $a_n = 0$ for $n < 0$ *i.e.* if the singularity in the variable q is removable. We also say that it is a *cusp form* if it is a modular form and $a_0 = 0$. Note that it is a modular form if and only if $\lim_{\Im z \to \infty} f(z) \in \mathbb{C}$, and in this case $g(q) = a_0 + O(q)$ and therefore $f(z)$ converges exponentially fast to a_0 as $\Im z \to \infty$. On the other hand if $a_{-1} \neq 0$ then $f(q)$ must be $\Omega(|q|^{-1})$ and therefore $|f(z)|$ cannot be bounded by a polynomial in $\Im z$. Some books use this as shortcut to define which

modular functions are modular forms: those which are bounded as $\Im z \to \infty$, or, as we will see in §2.6, those which grow at most polynomically fast as $\Im z \to 0^+$.

Note that by the functional equation (2.1) the values of a modular function are determined once we have specified the function on the fundamental domain $\mathbb{F}$. In the simplest case, when $k = 0$, the functions are invariant and therefore live in the quotient space $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$. This space can be visualized as the result of gluing together the sides of $\mathbb{F}$ as indicated by the transformations T and S. If we remove the orbits of the elliptic points i and ρ, the resulting quotient is both a riemannian manifold and a Riemann surface, and in particular there is a well-defined notion of angle. Around the image of the points i and ρ in the quotient, however, we can find neighbourhoods which do not have a full 2π circumference, as can be seen in figure 1.2. In the case of i the angle is just π, while in the case of ρ the angle is $2\pi/3$. These can be visualized as "cone"-like singularities in the quotient $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$. If we compactify the quotient by adding the missing point ∞, we also have a singularity around it, but in this case with a neighbourhood of zero radians around it. These three singularities can be resolved if we forget the metric and only care about the complex structure, by adding ad hoc charts which "multiply" the angles around them by the correct amount. These are similar to z^2 and z^3 for i and ρ and the exponential change of variables we introduced earlier for ∞. After this is done, we are left with a compact Riemann surface where analytic functions precisely correspond to modular forms of weight zero (see §2.2.5 of [**82**]). Of course we only have the constants by Liouville's theorem, but since $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$ is an interesting space from the number theoretic viewpoint (parametrizes elliptic curves over $\mathbb{C}$, for example) we want to construct non-trivial meromorphic functions over it. One way to do this is by quotienting modular forms, as the functional equation (2.1) shows.[1] This is one motivation to study this kind of functions, similar to the original motivation by Jacobi to study his theta function. A different, more pragmatic, motivation to study modular forms is simply that there are many examples of interesting functions arising from different contexts that satisfy (2.1), or generalizations of this equation. A third motivation is that for $k = 2$ the differential form $f(z)\,dz$ is invariant under the action of $\mathrm{SL}_2(\mathbb{Z})$, since a simple computation shows that $(\gamma z)' = (cz + d)^{-2}$, and therefore weight two modular forms provide holomorphic differential forms on the Riemann surface obtained by compactifying $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$. This is the reason they are called *modular forms*, the quotient space $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$ being called the *modular curve*. The differential forms thus obtained can then be integrated over paths on the modular curve, leading to the important theory of Eichler–Shimura [**28**].

A different point of view is the following: suppose we have a complex-valued function $g : \mathcal{L} \to \mathbb{C}$, where $\mathcal{L}$ is the space of lattices, and that it is $(-k)$-homogeneous in the sense that $g(\lambda\Lambda) = \lambda^{-k} g(\Lambda)$ for every $\lambda \in \mathbb{C}^*$ and every $\Lambda \in \mathcal{L}$. If we define $f(z) = g(\Lambda_z)$, where Λ_z is the lattice generated by 1 and $z \in \mathbb{H}$, then this function $f : \mathbb{H} \to \mathbb{C}$ captures all information about g. This is so because if $\alpha, \beta \in \mathbb{C}$ constitute a positively ordered based for Λ then $g(\Lambda) = g(\alpha\Lambda_{\beta/\alpha}) = \alpha^{-k} f(\beta/\alpha)$. Moreover we have the identity $\Lambda_z = (cz+d)\Lambda_{\gamma z}$, as the lattice on the right hand side is generated by $cz + d$ and $az + d$, which constitutes another basis for Λ_z. This is precisely (2.1) for f, and viceversa any $f : \mathbb{H} \to \mathbb{C}$ satisfying (2.1) can be translated to some

[1] In fact, since the resulting Riemann surface is conformally equivalent to the Riemann sphere, we know that the field of meromorphic functions can be generated by a single function. One such function is the so-called *j-invariant*, which appears naturally in the theory of elliptic curves precisely as an invariant of the isomorphism class of the curve (see §VII.3.3 of [**85**]).

k-homogeneous function g on the space of lattices. This provides a cheap way to construct examples, the simplest ones being *Eisenstein series* $E_k(\Lambda) = \sum_{0\neq\lambda\in\Lambda} \lambda^{-k}$, or in the z-variable $E_k(z) = \sum_{\vec{0}\neq(n,m)\in\mathbb{Z}^2}(nz+m)^{-k}$. This series absolutely converges to a nonzero modular function for every even integer $k \geq 4$, which is a modular form of weight k since $\lim_{\Im z\to\infty} E_k(z) = 2\zeta(k)$. Using the taylor series for the cotangent function it can be shown (see §VII.4 of [**85**]) that they admit the Fourier expansion

$$E_k(z) = 2\zeta(k) + \frac{2(2\pi i)^k}{(k-1)!}\sum_{n\geq 1}\sigma_{k-1}(n)q^n \tag{2.3}$$

where $q = e^{2\pi iz}$ and $\sigma_{k-1}(n) = \sum_{d|n} d^{k-1}$. In general it is often the case that the Fourier coefficients of modular forms are arithmetically interesting (and multiplicative!) functions.

Modular forms (or modular functions) of a fixed weight form a $\mathbb{C}$-vector space, as can be easily shown from the linearity of (2.1) and the rest of properties. We can also multiply modular forms (or functions) of weights k_1 and k_2 to obtain one of weight $k_1 + k_2$, and therefore if we take them all together they generate a graded $\mathbb{C}$-algebra. We can also divide them, but then we have to be careful to avoid introducing poles.

An important fact is that modular forms of a fixed weight are a finite-dimensional vector space over $\mathbb{C}$. A simple way to show the finiteness is by integrating the logarithmic derivative of a modular form around the boundary of the fundamental domain $\mathbb{F}$ and employing the functional equation (2.1) to find a very particular version of the Riemann-Roch formula, as done in §VII.3 of [**85**]. This formula provides strong restrictions on the functions that happen to be modular forms, and can be used to prove that the space of forms of weight k admits as a basis the set of all the products $E_4^n E_6^m$ for which $4n + 6m = k$, $n \geq 0$, $m \geq 0$ (corollary 2 of §VII of [**85**]). In particular there are only nonzero modular forms when $k \geq 0$ is an even integer, and in this case the space of modular forms has dimension $d = \lfloor k/12 \rfloor$ if $k \equiv 2 \pmod{12}$ and $d = \lfloor k/12 \rfloor + 1$ when $k \not\equiv 2 \pmod{12}$. Moreover the first d Fourier coefficients uniquely determine the modular form. Playing with these facts and the expansion (2.3) it is not hard to find surprising relations between certain Dirichlet convolutions of the functions σ_{k-1}.

The vector space of modular functions of a fixed weight, however, need not be of finite dimension. This is similar to what happens in other areas of mathematics, for example in PDEs. Consider as a model the heat equation on the real line. Once the initial conditions have been stablished one can only ensure uniqueness of the solution if one limits its growth at infinity. For modular forms, the equation is not differential but functional, and the initial conditions can be thought as prescribing a big enough but finite number of Fourier coefficients.

For weight 12 we have for the first time a nonzero cusp form, as the two modular forms E_4^3 and E_6^2 both have weight 12 and are linearly independent. Indeed, the combination $\Delta = 10800(20E_4^3 - 49E_6^2)$ is cuspidal, as is readily shown using the identities $\zeta(4) = \pi^4/90$ and $\zeta(6) = \pi^6/945$. The fact that Δ does not vanish identically can also be checked directly by computing the Fourier coefficient $\tau(1) = 1$ of the Fourier expansion $\Delta(z) = \sum_{n\geq 1}\tau(n)q^n$ from (2.3). The function Δ is called the *discriminant function*, while $\tau(n)$ is called *Ramanujan's tau function*. The latter was notably introduced by Ramanujan in 1916, who conjectured that it was multiplicative and for primes satisfied the bound $|\tau(p)| \leq 2p^{11/2}$. The multiplicativeness was

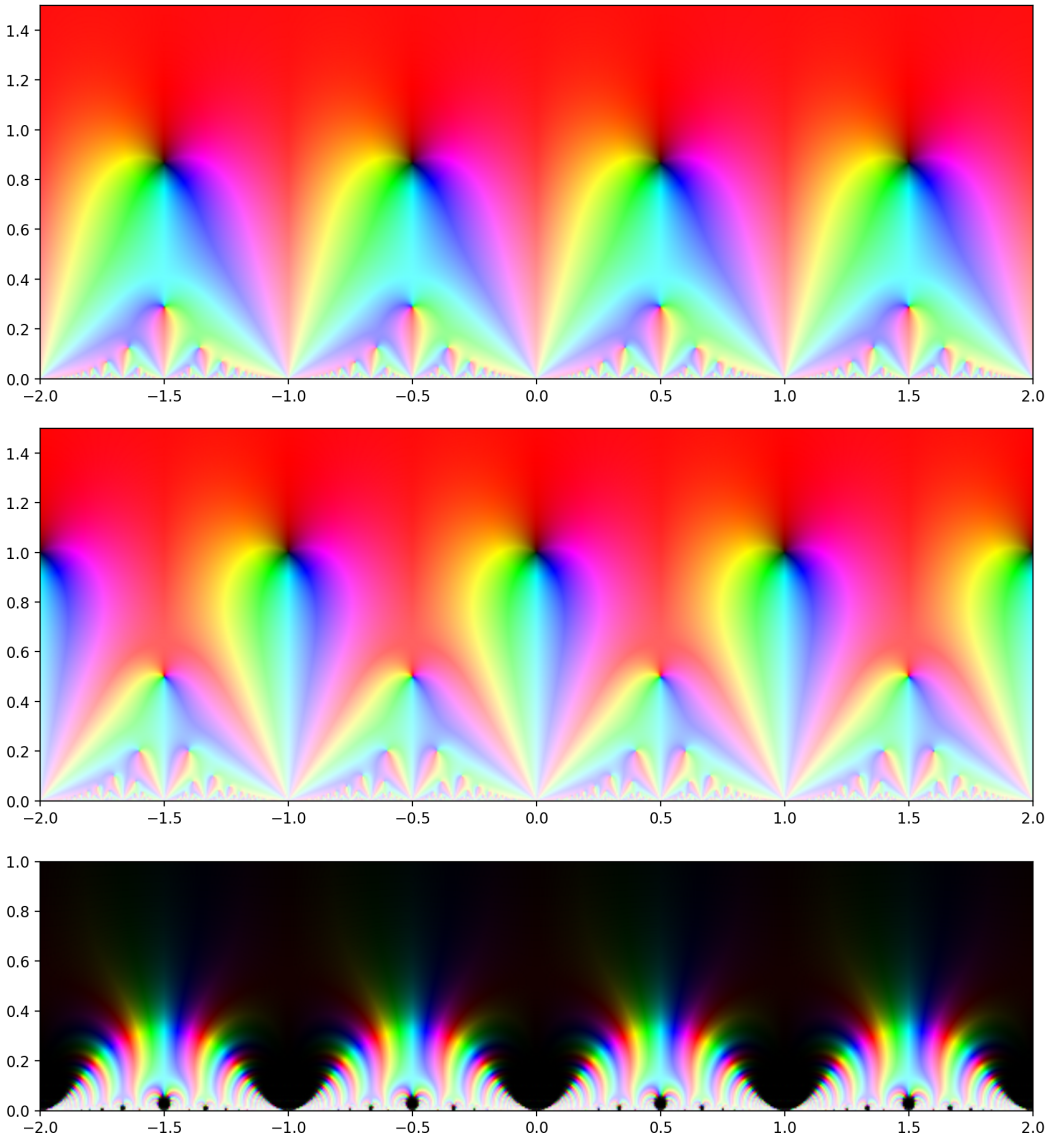


FIGURE 2.1. The modular forms E_4, E_6 and Δ (top to bottom). As is customary, in these plots a function $f : \mathbb{C} \to \mathbb{C}$ is represented by coloring the complex plane. The lightness of a point z indicates the modulus $|f(z)|$, where black means 0 and white ∞; while the hue indicates the argument of $f(z)$, the positive real numbers being red and the negative ones being cyan.

proven by Mordell one year later (nowadays usually shown with help of the Hecke operators, see §VII.5 of [**85**]), but the bound on its size would have to wait to 1974 when Deligne proved it as a consequence of Weil conjectures. Note the substantial difference with the coefficients of the Eisenstein series, which for primes are of the order p^{k-1}, where k is the weight. Indeed, the coefficients of cusp forms are always much smaller than the coefficients of non-cuspidal forms, as we will show in §2.6, and sharp bounds are usually very deep results (*cf.* [**25**]).

The examples we have given, and in particular the Eisenstein series E_4 and E_6 and the discriminant function Δ, appear naturally in the theory of elliptic curves

over $\mathbb{C}$. The first two, when evaluated at a lattice Λ, provide the coefficients of the Weierstrass form for the elliptic curve $\mathbb{C}/\Lambda$, while the discriminant function coincides with the discriminant of such polynomial (see §VII.2 of [**85**]). The j-invariant is — as the name suggest— an invariant of the isomorphism class of the elliptic curve. A plot of these functions is included in figure 2.1, where it is apparent how the functional equation (2.1) relates their values among different Ford circles.

2.2. Multiplier systems

The vast majority of modular forms which one encounters "in the wild" do not conform to the definition we have just given. This is for example the case of Jacobi's theta function θ, defined in (**I**.1). For this function, (2.2) has to be replaced with

$$\begin{cases} \theta(z+2) = \theta(z), \\ \theta(-1/z) = \sqrt{-iz}\,\theta(z). \end{cases} \tag{2.4}$$

The first equation is clearly satisfied by definition. To see the second holds we use the fact that the gaussian $f(x) = e^{-\pi x^2}$ is its own Fourier transform, and therefore for $g(x) = f(x\sqrt{t})$ we have $\hat{g}(\xi) = t^{-1/2} f(x/\sqrt{t})$. Applying Poisson's summation formula to g we obtain the second equation for $z = it$, and the identity principle shows it must hold for any $z \in \mathbb{H}$.

Note (2.4) almost mimics (2.2) for weight $k = 1/2$. We have however to accomodate two facts: firstly the group of transformations —generated in this case by T^2 and S— is a finite index proper subgroup of $\mathrm{SL}_2(\mathbb{Z})$. Secondly, there is an unimodular factor $\sqrt{-i}$ multiplying the right hand side of the second equation. When applying repeatedly these two equations we obtain a general functional equation similar to (2.1) but with a different unimodular constant μ_γ for every transformation γ in the transformation group.

Take now any finite index subgroup Γ of $\mathrm{SL}_2(\mathbb{Z})$. We are interested in nonzero functions $f : \mathbb{H} \to \mathbb{C}$ satisfying for some $k \in \mathbb{R}$,

$$f(\gamma z) = \mu_\gamma (cz+d)^k f(z) \qquad \text{for every} \quad \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma, \tag{2.5}$$

where μ_γ is an unimodular constant depending on γ. The power function, and any logarithm we consider, will always correspond to the principal branch, with argument determination in $(-\pi, \pi]$. Note we may assume without loss of generality that $-1 \in \Gamma$, as otherwise we may simply add it to the subgroup and appropriately choose $\mu_{-\gamma}$ for every $\gamma \in \Gamma$ to make (2.5) hold for the new group. Once this is done, the functional equations corresponding to γ and $-\gamma$ are redundant (in fact it is the group $\Gamma/\{\pm 1\}$ the one acting), hence for the sake of simplicity we will often take γ satisfying the following convention:

$$c > 0, \text{ or } c = 0 \text{ and } d > 0, \text{ where } (c,d) \text{ is the bottom row of } \gamma. \tag{2.6}$$

Note these matrices do not form a group, simply a transversal set for $\Gamma/\{\pm 1\}$.

For convenience we will also use the notation $j_\gamma(z) = cz + d$. Note if $z \in \mathbb{H}$ then (2.6) is equivalent to $\arg j_\gamma(z) \in (\pi, 0]$. The function j_γ also satisfies the following properties:

Proposition 2.1. *For any $\gamma, \eta \in \mathrm{SL}_2(\mathbb{R})$ and $z, w \in \mathbb{H}$ we have:*

(*i*) $j_{\gamma\eta}(z) = j_\gamma(\eta z) j_\eta(z)$.

(*ii*) $j_{\gamma^{-1}}(z) = \big(j_\gamma(\gamma^{-1}z)\big)^{-1}$.

(*iii*) $(\gamma w - z)j_\gamma(w) = (w - \gamma^{-1}z)j_{\gamma^{-1}}(z)$.

(*iv*) *For any fixed* $k \in \mathbb{R}$, *the following expression does not depend on* z:

$$c(\gamma,\eta) = \frac{\big(j_\gamma(\eta z)\big)^k \big(j_\eta(z)\big)^k}{\big(j_{\gamma\eta}(z)\big)^k}.$$

Proof. The first property can be easily checked by substitution. Choosing $\eta = \gamma^{-1}$ we obtain the second one. The third is equivalent using (*ii*) to $w - u = (\gamma w - \gamma u)j_\gamma(w)j_\gamma(u)$, where $u = \gamma^{-1}z$, identity which can also be checked by substitution.

To show (*iv*) note that if u, v are complex numbers the quantity $u^k v^k (uv)^{-k}$ depends only on the unique integer n for which $\arg u + \arg v - \arg(uv) = 2\pi n$. Hence if $\arg u$, $\arg v$ and $\arg uv$ all vary continuously, the expression $u^k v^k (uv)^{-k}$ must remain constant. Since in our case $u = j_\gamma(\eta z)$, $v = j_\eta(z)$ and $uv = j_{\gamma\eta}(z)$ due to (*i*), it suffices to show that $\arg j_\sigma(z)$ is a continuous function of z for $z \in \mathbb{H}$ and for any $\sigma \in \mathrm{SL}_2(\mathbb{R})$. This is a consequence of the fact that, depending on the sign of the bottom row of σ, $j_\sigma(z)$ varies continuously and remains in one of the following four regions: $\mathbb{H}$, $\mathbb{R}^+$, $\mathbb{R}^-$ or $-\mathbb{H}$. □

The first property together with the associative law $\gamma(\eta z) = (\gamma\eta)z$ imply that if we want a nonzero function f to satisfy (2.5) then the constants μ must satisfy for any $\gamma, \eta \in \Gamma$ the identity

$$\mu_{\gamma\eta} = c(\gamma,\eta)\,\mu_\gamma \mu_\eta, \tag{2.7}$$

where the constant is given by (*iv*) of proposition 2.1. Any function $\gamma \mapsto \mu_\gamma$ on Γ satisfying $|\mu_\gamma| = 1$, $\mu_{-1} = e(-k/2)$ (recall $e(x) = e^{2\pi i x}$) and (2.7) for all $\gamma\eta \in \Gamma$ is called a *multiplier system of weight* k *on* Γ. If $\mu_\gamma = 1$ for every $\gamma \in \Gamma$ the multiplier system is said to be trivial. Note also when the weight is an integer the multiplier system is simply an homomorphism from Γ to S^1 sending the matrix -1 to $(-1)^k$. For more information on multiplier systems we refer the reader to §3 of [**82**].

Once we have a multiplier system μ we do not have any obvious obstructions to (2.5), in the sense that we may always construct nonzero continuous functions $f : \mathbb{H} \to \mathbb{C}$ satisfying this functional equation. If f is also holomorphic we say it is a *modular function of weight* k *for* Γ *and multiplier system* μ. Nonzero modular functions can always be constructed for any weight $k > 2$ generalizing the Eisenstein series described in the previous section (see §5.1 of [**82**]). On the other hand, if we directly find a nonzero function satisfying (2.5) for some unimodular constants μ_γ we can automatically guarantee they provide a multiplier system. In particular, for Jacobi's theta funcion θ the transformation group $\Gamma_\theta = \langle T^2, S, -1\rangle$, sometimes called the *theta group*, can be characterized as the set of matrices in $\mathrm{SL}_2(\mathbb{Z})$ of the form $\left(\begin{smallmatrix} \text{odd} & \text{even} \\ \text{even} & \text{odd} \end{smallmatrix}\right)$ or $\left(\begin{smallmatrix} \text{even} & \text{odd} \\ \text{odd} & \text{even} \end{smallmatrix}\right)$, and the multiplier system is determined by $\mu_\gamma = 1$ if $c = 0$ and the incomplete Gaussian sum

$$\mu_\gamma = \left(\sqrt{\frac{i}{c}} \sum_{j=0}^{c-1} e\big(-dj^2/(2c)\big)\right)^{-1} \qquad \text{if } c > 0.$$

A simple proof of these facts is provided by Duistermaat in §3 of [**24**]. The multiplier is always an eighth root of the unity as it follows by completing and evaluating the Gauss sum, or directly from the fact that (2.5) for an arbitrary $\gamma \in \Gamma_\theta$ must be obtained by adequately composing the identites (2.4).

An alternative and sometimes more convenient way of writing (2.5) involves the slash operator. Given any $\gamma \in \mathrm{GL}_2(\mathbb{R})$ with positive determinant we define the slash operator $|_\gamma$ of weight k acting on the functions $f : \mathbb{H} \to \mathbb{C}$ in the following way:

$$f|_\gamma(z) = \big(\det \gamma\big)^{k/2} \frac{f(\gamma z)}{\big(j_\gamma(z)\big)^k}.$$

It depends on the weight k, but this dependence is usually omitted as k is fixed. The slash operator satisfies the composition law $(f|_\gamma)|_\eta = c(\gamma,\eta) f|_{\gamma\eta}$, as can be readily checked by substituting.

Using the slash operator, (2.5) admits the compact form

$$f|_\gamma = \mu_\gamma f \qquad \text{for any } \gamma \in \Gamma.$$

The following proposition describes what happens when $\gamma \notin \Gamma$.

PROPOSITION 2.2. *Suppose $f : \mathbb{H} \to \mathbb{C}$ is a nonzero modular function of weight k for the subgroup Γ and multiplier system μ, and take $\gamma \in \mathrm{GL}_2^+(\mathbb{R})$. Then there is a multiplier system ν of weight k for the group $\Gamma' = \gamma^{-1}\Gamma\gamma \cap \mathrm{SL}_2(\mathbb{Z})$ such that $f|_\gamma$ is a modular function of weight k for Γ' and multiplier system ν.*

PROOF. The function $f|_\gamma$ is clearly holomorphic on $\mathbb{H}$, as $j_\gamma(z)$ never crosses the branch of w^k. If we take $\eta = \gamma^{-1}\sigma\gamma$ with $\sigma \in \Gamma$, the composition law for the slash operator implies

$$(f|_\gamma)|_\eta = c(\gamma,\eta)\, f|_{\sigma\gamma} = c(\gamma,\eta)\, c(\sigma,\gamma)^{-1}\, (f|_\sigma)|_\gamma,$$

and since f is a modular function for Γ,

$$(f|_\gamma)|_\eta = \mu_\sigma\, c(\gamma,\eta)\, c(\sigma,\gamma)^{-1} f|_\gamma.$$

Since the constant $\nu_\eta = \mu_\sigma\, c(\gamma,\eta)\, c(\sigma,\gamma)^{-1}$ is unimodular and $f|_\gamma$ is nonzero, this shows at once that ν is a multiplier system for Γ' and $f|_\gamma$ a modular function for Γ' and ν. □

2.3. The action of finite order subgroups

Let Γ be a finite index subgroup of $\mathrm{SL}_2(\mathbb{Z})$ and fix a set of representatives $\eta_1, \ldots, \eta_n$ of the right cosets of Γ, where n is the index $[\mathrm{SL}_2(\mathbb{Z}) : \Gamma]$. The union $\mathbb{F}_\Gamma = \cup_j \eta_j \mathbb{F}$ is always a fundamental domain for Γ.[2] Indeed, the translates $\{\gamma\mathbb{F}_\Gamma\}_{\gamma\in\Gamma}$ cover the upper half-plane because $\mathrm{SL}_2(\mathbb{Z})$ decomposes as $\cup_j \Gamma\eta_j$, while two translates can never intersect at an interior point because otherwise two translates of $\mathbb{F}$ would also do. In fact, we can always choose the right-transversal $\eta_1, \ldots, \eta_n$ so that both $\mathbb{F}_\Gamma$ and its interior are connected sets. This is a consequence of the following property: let Ω be an union of translates of $\mathbb{F}$ satisfying that any translate of $\mathbb{F}$ sharing an edge with Ω is related modulo Γ to some translate in Ω. Then $\mathbb{H} \subset \cup_{\gamma\in\Gamma}\gamma\Omega$, as we can use elements of Γ to translate Ω and cover any translate of $\mathbb{F}$ sharing an edge with Ω, and then again to cover any translate sharing an edge with the new set, and recursively fill up the whole upper half-plane.

Suppose now that Ω is a connected component of $\mathbb{F}_\Gamma$. If Ω is a proper subset of $\mathbb{F}_\Gamma$ then some translate of $\mathbb{F}$ with an edge in common with Ω must be related modulo Γ to some $\eta_j\mathbb{F}$ in a different component of $\mathbb{F}_\Gamma$. We may therefore adjust η_j to move $\eta_j\mathbb{F}$ so it forms part of the connected component Ω, and repeat the procedure.

[2] In contrast, the set $\mathbb{F}'_\Gamma = \cup_j \eta_j \mathbb{F}'$ is not always a strict fundamental domain because it may contain some points in the orbits of i and ρ modulo $\mathrm{SL}_2(\mathbb{Z})$ which are related modulo Γ.

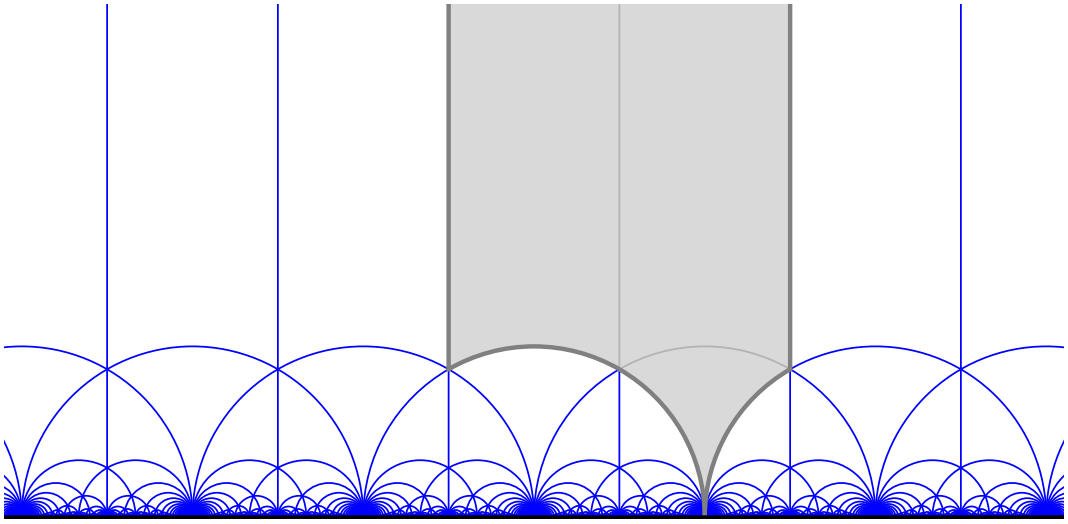


FIGURE 2.2. A fundamental domain for the group Γ_θ, comprising the translates $\mathbb{F}$, $T\mathbb{F}$ and $TS\mathbb{F}$. It has as limit points the cusps ∞ and 1.

The aforementioned property also shows that Γ is finitely generated: if $\gamma_1, \ldots, \gamma_r \in \Gamma$ are chosen so that $\cup_j \gamma_j \mathbb{F}_\Gamma$ covers all the translates of $\mathbb{F}$ sharing an edge with $\mathbb{F}_\Gamma$, then immediately $\mathbb{F}_\Gamma$ is also a fundamental domain for the subgroup generated by the γ_j and -1. But since the translates $\gamma\mathbb{F}_\Gamma$ have disjoint interiors, this group must necessarily coincide with Γ.

These simple ideas are a powerful tool to find fundamental domains. For the theta group Γ_θ, for example, they easily lead to the fundamental domain $\mathbb{F} \cup T\mathbb{F} \cup TS\mathbb{F}$, shown in figure 2.2.

When we let Γ act on the set of cusps $\mathbb{Q} \cup \{\infty\}$ the unique orbit modulo $\mathrm{SL}_2(\mathbb{Z})$ also breaks into finitely many orbits, but here we cannot guarantee the number of orbits to equal the index of the group. The set $\{\eta_1\infty, \ldots, \eta_n\infty\}$ always contains a point in every orbit, but some orbits may contain more than one point. If x is a cusp, its equivalence class will be denoted $[x]$ when there is no ambiguity on which is the group acting.[3] The stabilizer of x in $\mathrm{SL}_2(\mathbb{Z})$ is a subgroup isomorphic to $\mathbb{Z}$, and under this isomorphism the stabilizer of x in Γ corresponds to some subgroup $m_x\mathbb{Z}$ of index $m_x \geq 1$. The positive integer m_x is called the *width* of the cusp x (with respect to Γ). Two cusps in the same orbit modulo Γ have conjugated stabilizers and therefore the same width, hence $m_{[x]}$ is well defined.

The width m_x coincides with the number of translates $\eta_j\mathbb{F}$ in $\mathbb{F}_\Gamma$ whose missing vertex lies in the orbit $[x]$. We show this for $x = \infty$, while for other cusps is similar. On the one hand, the width m_∞ is the minimum positive integer m such that $T^m \in \Gamma$. On the other hand, we can change the η_j so that all the $\eta_j\mathbb{F}$ having the missing vertex in $[\infty]$ actually have ∞ as the missing vertex, and then again so that they are of the form $T^j\mathbb{F}$ for $0 \leq j < m_\infty$. Now, if there is a missing spot, we must be able to fill it by translating some $\eta_{j_0}\mathbb{F}$ by some $\gamma \in \Gamma$. But then the missing vertex of this $\eta_{j_0}\mathbb{F}$ must lie in $[\infty]$ and this implies $\gamma = T^m$ for some $0 < m < m_\infty$, contradicting the choice of m_∞.

As a consequence the index of Γ coincides with the sum of the widths of all the orbits of cusps modulo Γ, *i.e.* $[\mathrm{SL}_2(\mathbb{Z}) : \Gamma] = \sum m_{[x]}$. The number of equivalence

[3] Some authors use the term cusp to refer to the equivalence class instead of to the point.

classes of cusps coincides with the "missing" points of the surface $\Gamma\backslash\mathbb{H}$ which we must add to compactify it. The compactified quotient again admits a structure of Riemann surface after removing the singularities introduced by the elliptic transformations of Γ and the added cusps.

All the equivalence classes of cusps modulo Γ are dense in $\mathbb{R}$. In fact, we have the following stronger result:

PROPOSITION 2.3. *Let Γ be a finite index subgroup of $\mathrm{SL}_2(\mathbb{Z})$, α an irrational number and $x \in \mathbb{Q} \cup \{\infty\}$. Then there are infinitely many rationals $p/q \in [x]$ satisfying*

$$\left|\alpha - \frac{p}{q}\right| \leq \frac{C}{q^2}$$

for some constant $C > 0$ depending only on the group Γ.

PROOF. The vertical ray $\{\Re z = x\}$ cuts the boundary of infinitely many generalized Ford circles for $\delta = 2$ at a sequence of points $z_n = \alpha + iy_n$ where $y_n \to 0^+$, and for every n we can find some $\eta \in \mathrm{SL}_2(\mathbb{Z})$ such that $\eta(z_n)$ lies in the segment $I = \{\Im z = 1/2, -m_\infty/2 \leq \Re z \leq m_\infty/2\}$. This is so because we can transform the Ford circle where z_n lies to $\mathcal{F}_\infty(2)$ and then compose with a translation if necessary. Decomposing $\eta^{-1} = \gamma\eta_i$ for some i and $\gamma \in \Gamma$, we have $\gamma^{-1}z_n \in \eta_i I$. Since the set $\cup_i \eta_i I$ is compact, for some C big enough the Speiser circle $\mathcal{F}_x(C)$ contains this union, and therefore $z_n \in \mathcal{F}_{\gamma x}(C)$. This implies the inequality we were looking for, for $p/q = \gamma x$. As there are finitely many equivalence classes of cusps the constant C can be taken to be uniform. □

2.4. Expansion at the cusps

Let $f : \mathbb{H} \to \mathbb{C}$ be a modular function with respect to Γ and multiplier system μ. Let m_∞ be the width of ∞. Then $T^{m_\infty} \in \Gamma$ and the functional equation (2.5) reads $f(z+m_\infty) = e(\kappa_\infty)f(z)$ where $e(\kappa_\infty) = \mu_{T^{m_\infty}}$. If we define $g(z) = f(m_\infty z)e(-\kappa_\infty z)$ then g is holomorphic and 1-periodic, and therefore admits a Fourier expansion $g(z) = \sum_{n=-\infty}^{\infty} a_n e^{2\pi i n z}$. Translating this back to f,

$$f(z) = \sum_{n=-\infty}^{\infty} a_n e^{2\pi i (n+\kappa_\infty)z/m_\infty}. \tag{2.8}$$

THEOREM 2.4 (EXPANSION AT THE CUSPS). *Given $x \in \mathbb{Q} \cup \{\infty\}$ and $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ such that $\gamma x = \infty$, the modular function f admits the expansion*

$$f(z) = \big(j_\gamma(z)\big)^{-k} \sum_{n=-\infty}^{\infty} a_n\, e^{2\pi i (n+\kappa_x)\gamma z/m_x},$$

where m_x is the width of the cusp x and $0 \leq \kappa_x < 1$, both depending only on the class $[x]$. The modulus of the coefficients $|a_m|$ also depends only on the class $[x]$. Moreover, when $\kappa_x = 0$ the coefficient a_0 only depends on x, as long as γ is chosen satisfying (2.6).

PROOF. The expansion follows at once from (2.8) applied to the function $f|_{\gamma^{-1}}$, which is modular by proposition 2.2. Note m_∞ for the conjugated group is m_x for Γ. If γ_1 is another matrix for which $\gamma_1 x = \infty$ then $\gamma_1 = \pm T^m \gamma$ and by the uniqueness of the Fourier expansion the κ_x must coincide. The a_n must also vary by an unimodular constant, equal to $e(m(n+\kappa_x)/m_x)(\pm j_\gamma(z))^{-k}(j_\gamma(z))^k$. This constant is 1 if $m = \kappa_x = 0$ and the sign is positive. Finally if $x' = \eta x$ for some

$\eta \in \Gamma$ then $\gamma\eta^{-1}x' = \infty$ and when we use this matrix to compute the expansion we have $f|_{\eta\gamma^{-1}} = c(\eta, \gamma^{-1})^{-1}\mu_\eta f|_{\gamma^{-1}}$, *i.e.* we obtain the same expansion up to the unimodular constant $c(\eta, \gamma^{-1})^{-1}\mu_\eta$. □

In the proof we have applied the slash operator $|_{\gamma^{-1}}$ to the function f. The resulting function is essentially $f(\gamma^{-1}z)$, with the extra factor $\big(j_\gamma(z)\big)^{-k}$ included to keep the automorphy. Since $\gamma^{-1}z$ approaches x within Speiser circles when $\Im z \to \infty$, we are effectively moving the cusp x to infinity to have a "better look" at how f behaves close to x. Note that if two cusps belong to the same class modulo Γ the functional equation guarantees that f behaves in a similar way at both of them, and this is reflected in the statement of the theorem. Because of this there is essentially only one expansion per class of cusps, which can be made unique by fixing one particular choice of γ^{-1} for each of them.

Most authors also remove the width of the cusp in the expansion by hiding the change of variables $z \mapsto m_x z$ inside the linear fractional transformation appearing in the exponent. This is done by expanding $f|_{\gamma^{-1}\eta}$ instead of $f|_{\gamma^{-1}}$, where $\eta = \left(\begin{smallmatrix} \sqrt{m_x} & 0 \\ 0 & 1/\sqrt{m_x} \end{smallmatrix}\right)$. The matrix $\gamma^{-1}\eta$ usually receives the name of *scaling matrix*, and, as mentioned above, fixing one choice of scaling matrix per class of cusps suffices to make the Fourier expansion unique at every cusp. In this document we have preferred to avoid the use of scaling matrices altogether; instead in §2.5 we will show that under reasonable hypotheses we can rely on the trick of scaling f directly to avoid having to keep track of the cusp width at infinity.

We say that f is a *modular form of weight*[4] k *for the group* Γ *and multiplier system* μ if it is a modular function, and in the expansion provided by theorem 2.4 for every $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ we always obtain a Fourier series with only non-negative frequencies, *i.e.* $n + \kappa_x < 0$ implies $a_n = 0$. Note it suffices to check this only once for every orbit of cusps modulo Γ. Given a cusp x we define[5] $f(x)$ as the coefficient a_0 if $\kappa_x = 0$ and γ is chosen satisfying (2.6), or as 0 if $\kappa_x > 0$. If $f(x) = 0$ we say that f is cuspidal at x, or for convenience (although this is nonstandard) that x is cuspidal for f. This property, again, only depends on the orbit of x. If f is cuspidal at every cusp then we say that f is a *cusp form*.

When f is a modular form the expansion (2.8) converges absolutely and uniformly over the Speiser circles $\mathcal{F}_\infty(\delta)$ for $\delta > 0$, and exponentially fast as $\delta \to 0^+$. Since the action of $\mathrm{SL}_2(\mathbb{Z})$ preserves them, it is clear that the expansion provided by theorem 2.4 converges absolutely and uniformly over $\mathcal{F}_x(\delta)$ for $\delta > 0$. This provides a very precise approximation in these circles for small δ by truncating the Fourier series:

Corollary 2.5. *Let f be a modular form of weight k. Let p, q be coprime integers satisfying either $q > 0$ or $q = 0$ and $p = -1$, and fix $\delta_0 > 0$. Then, as long as*

[4]The weight is uniquely determined by theorem 2.4 for any nonzero f, for example by taking $\gamma = S$ and $z = i/t$ and considering the growth of f as $t \to \infty$.

[5]Had we decided to use scaling matrices, by definition of the slash operator all Fourier coefficients would also be multiplied by $m_x^{k/2}$. This different normalization, albeit unimportant, is also common in the literature, and one should be aware of this when comparing results from different sources. In particular it was used by the author in [**80**], leading to some minor differences between the proofs of chapter 3 and those included in the article.

$z \in \mathcal{F}_{p/q}(\delta_0)$, we have

$$f(z) = \frac{f(p/q)}{(qz-p)^k} + O\left((\Im z)^{-k/2} e^{-K\Im z|qz-p|^{-2}}\right),$$

where the constant $K > 0$ and the O-constant depend only on f and δ_0.

PROOF. Apply theorem 2.4 with some $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ whose lower row is $(q, -p)$ and $x = p/q$, and use (1.2) to obtain the bound

$$\left| f(z) - \frac{f(p/q)}{(qz-p)^k} \right| \leq (\Im z)^{-k/2} \, t^{k/2} g_x(t),$$

where

$$g_x(t) = \sum_{n+\kappa_x > 0} |a_n| \, e^{-2\pi(n+\kappa_x)t/m_x}$$

and $t = \Im z|qz-p|^{-2}$. Since the condition $z \in \mathbb{F}_x(\delta)$ is equivalent to $t \geq \delta^{-1}$ by lemma 1.1, the absolute and uniform convergence of the expansion at the cusp for f implies uniform convergence for the series defining g_x in the sets $t \geq \delta^{-1}$. In particular, $g_x(t) \leq C$ for $t \geq \delta_0^{-1}/2$. If we let $K' = \pi\kappa_x/m_x$ if $\kappa_x > 0$ and $K' = \pi/m_x$ when $\kappa_x = 0$, for any $t \geq \delta_0^{-1}$,

$$g_x(t) = g_x(t/2 + t/2) \leq Ce^{-K't}.$$

For any $0 < K < K'$ we therefore have $t^{k/2} g_x(t) \ll e^{-Kt}$, which is the bound we were looking for. The uniformity of the constants follows from the fract that the function g_x only depends on the equivalence class of the orbit of x modulo Γ, and therefore there are finitely many possibilities. □

The $\mathbb{C}$-vector space of modular forms of weight k for a finite index subgroup and a multiplier system is always finite-dimensional. As in the simplest case of forms for the whole modular group and trivial multiplier system, this follows from a version of Riemann-Roch for the compactification of the Riemann surface $\Gamma\backslash\mathbb{H}$, which can be proved by integrating the logarithmic derivative of a modular form on the boundary of every translate of $\mathbb{F}$ in the fundamental domain $\mathbb{F}_\Gamma$ (see §4.2 of [**82**]). When taken all the forms for the same group together they generate a graded $\mathbb{C}$-algebra, as the expansion at the cusps converges uniformly and can be termwise multiplied.

A new feature is that now the slash operator also relates them across different subgroups:

THEOREM 2.6. *Suppose $f : \mathbb{H} \to \mathbb{C}$ is a nonzero modular form of weight k for the finite index subgroup Γ and multiplier system μ, and take $\gamma \in \mathrm{GL}_2^+(\mathbb{R})$ satisfying that the subgroup $\Gamma' = \gamma^{-1}\Gamma\gamma \cap \mathrm{SL}_2(\mathbb{Z})$ is of finite index. Then there is a multiplier system ν of weight k for Γ' such that $f|_\gamma$ is a modular form of weight k for Γ' and multiplier system ν.*

PROOF. By proposition 2.2 the function $f|_\gamma$ is a modular function. To see it is a modular form it suffices to see that for any $\eta \in \mathrm{SL}_2(\mathbb{Z})$ the limit $\lim_{\Im z \to \infty} (f|_\gamma)|_\eta(z)$ exists. This follows from the composition law for the slash operator and the fact that f is a modular form. □

As a consequence every modular form f always has some "companions" $f|_{\eta_1}, \ldots, f|_{\eta_r}$ where the $\eta_i \in \mathrm{SL}_2(\mathbb{Z})$ are chosen satisfying $\eta_i \infty = x_i$ for a set of representatives $\{x_i\}$ of the different orbits of cusps modulo Γ distinct from $[\infty]$ (corresponding to

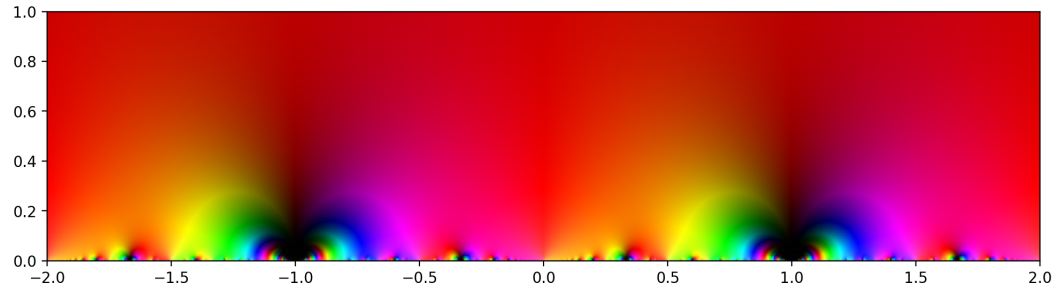


FIGURE 2.3. The modular form θ.

a choice of the scaling matrices). These modular forms, together with f itself, are essentially the "different views" of f at different cusps.

We go back to the example of Jacobi's theta function. Since the theta group Γ_θ has only two equivalence classes of cusps with representants ∞ and 1 (see figure 2.2) to see that Jacobi's theta function is a modular form we have only to check $\lim_{\Im z\to\infty} \theta|_\eta(z) \in \mathbb{C}$ with $\eta(\infty) = \infty$ and $\eta(\infty) = 1$. In the first case, $\lim_{\Im z\to\infty} \theta(z) = 1$ follows from the definition. In the second case, however, we require an explicit expression for $\theta|_{TS}$. At the same time, since $1, T, TS$ must be right-transversal for Γ_θ, we can also compute, up to constant, all the possible functions $\theta|_\eta$:

PROPOSITION 2.7. *Let $q = e^{\pi i z}$. We have for $\theta(z) = \sum_{n\in\mathbb{Z}} q^{n^2}$ the identities*

$$\theta|_T(z) = \sum_{n\in\mathbb{Z}} (-1)^n q^{n^2} \qquad \text{and} \qquad \theta|_{TS}(z) = \sqrt{-i} \sum_{n\in\mathbb{Z}} q^{(n+1/2)^2}.$$

PROOF. The first identity is evident. For the second one we apply Poisson summation to the function $f(x) = e^{-\pi(x+1/2)^2 t}$. Using that $g(x) = e^{-\pi x^2}$ is its own Fourier transform and elementary properties, we have $\hat{f}(\xi) = t^{-1/2} e^{\pi i \xi} e^{-\pi \xi^2/t}$. Poisson summation then shows $\theta|_T(-1/z) = \sqrt{-iz} \sum_{n\in\mathbb{Z}} q^{(n+1/2)^2}$ for $z = it$. □

Hence $\lim_{\Im z\to\infty} \theta|_{TS}(z) = 0$ and θ is cuspidal at [1] (this is clear in figure 2.3). The Fourier expansion also shows $\kappa_1 = 1/2$.

2.5. Congruence subgroups

In many treatises Jacobi's theta function is defined as $\sum_{n\in\mathbb{Z}} e(n^2 z)$ instead, *i.e.* as $\theta(2z)$ with our notation. The reason is that this is also a modular form, as up to a constant it equals $\theta|_\sigma$ where $\sigma = \left(\begin{smallmatrix} \sqrt{2} & 0 \\ 0 & 1/\sqrt{2} \end{smallmatrix}\right)$, as it can be checked that $\sigma^{-1}\Gamma\sigma \cap \mathrm{SL}_2(\mathbb{Z})$ is a finite index subgroup. The advantage is that it is given by a Fourier series with only integer frequencies, the disadvantage is that now the group of simmetries is smaller, it has more equivalence classes of cusps and these are wider. Nevertheless it will be useful to be able to reduce to such Fourier series to simplify the forthcoming results.

To be able to do this for arbitrary modular forms we need the following two conditions to hold:

(*i*) The group $\sigma_m^{-1}\Gamma\sigma_m \cap \mathrm{SL}_2(\mathbb{Z})$ is a subgroup of finite index of $\mathrm{SL}_2(\mathbb{Z})$ for every integer m, where σ_m is the *scaling matrix* $\left(\begin{smallmatrix} \sqrt{m} & 0 \\ 0 & 1/\sqrt{m} \end{smallmatrix}\right)$.

(*ii*) For every cusp $x \in \mathbb{Q} \cup \{\infty\}$ the parameter κ_x is a rational number.

Condition (ii) also admits an equivalent formulation: using the definition of the multiplier system for $f|_\gamma$ provided in the proof of proposition 2.2 it can be seen that $e(\kappa_x) = \mu_\eta$ where $\eta = \gamma^{-1}T^{m_x}\gamma$ and $\gamma x = \infty$. Moreover $\mu_{\eta^n} = \mu_\eta^n$ for any integer n and these matrices always have trace $+2$. Hence (ii) is satisfied if and only if μ_η is a root of unity for any parabolic $\eta \in \Gamma$ of positive trace.

Given any modular form f for Γ and μ satisfying the above properties, for an appropriately chosen m the function $f|_{\sigma_m}$ is a again a modular form and has a Fourier expansion at ∞ (2.8) with only integer frequencies. Note that not necessarily $m_\infty = 1$, simply $a_n = 0$ when $m_\infty \nmid n$.

As we are going to show, condition (i) is automatically satisfied by any finite index subgroup of $\mathrm{SL}_2(\mathbb{Z})$, so it imposes no new restriction. Condition (ii) however needs to be checked; an example of this is given in §6.4 of [**82**]. We are going to assume it is satisfied by any multiplier system considered in the rest of this dissertation, although all the results can however be extended to remove this hypothesis with little effort if needed.

To show all finite index subgroups satisfy (i) we need to introduce an important class of subgroups, the *congruence subgroups*. The *principal congruence subgroup of order* N, denoted by $\Gamma(N)$, is the subgroup composed of all those matrices (entry-wise) congruent to the identity modulo N. A *congruence subgroup* is any subgroup containing $\Gamma(N)$ for some N. As $\Gamma(N)$ is a normal subgroup, being given by the kernel of the homomorphism $\mathrm{SL}_2(\mathbb{Z}) \to \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$, all congruence subgroups containing $\Gamma(N)$ may be identified with subgroups of $\mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$. As a consequence we can characterize the congruence subgroups as those that can be described by a finite number congruences modulo some N. The minimum N for which $\Gamma(N)$ is contained in Γ is called the *level of* Γ. In particular the theta group Γ_θ is a congruence subgroup of level 2.

If Γ is a congruence subgroup of level N then $\sigma_m^{-1}\Gamma\sigma_m \cap \mathrm{SL}_2(\mathbb{Z})$ is always a congruence subgroup of level at most mN. To see this note that

$$\sigma_m \begin{pmatrix} a & b \\ c & d \end{pmatrix} \sigma_m^{-1} = \begin{pmatrix} a & bm \\ c/m & d \end{pmatrix}.$$

Hence if $\gamma \in \Gamma(mN)$ then $\sigma_m\gamma\sigma_m^{-1} \in \Gamma(N)$ or $\gamma \in \sigma_m^{-1}\Gamma(N)\sigma_m$. In particular $\Gamma(m)$ is always contained in $\sigma_m^{-1}\,\mathrm{SL}_2(\mathbb{Z})\sigma_m$. For an arbitrary finite index subgroup Γ of $\mathrm{SL}_2(\mathbb{Z})$ we have

$$[\Gamma(m) : \sigma_m^{-1}\Gamma\sigma_m \cap \Gamma(m)] \leq [\sigma_m^{-1}\,\mathrm{SL}_2(\mathbb{Z})\sigma_m : \sigma_m^{-1}\Gamma\sigma_m] < \infty.$$

Hence $\sigma_m^{-1}\Gamma\sigma_m \cap \mathrm{SL}_2(\mathbb{Z})$ must also be of finite index in $\mathrm{SL}_2(\mathbb{Z})$, which shows (i) always holds.

A family of arithmetically relevant congruence subgroups are the *Hecke congruence subgroups* $\Gamma_0(N)$, defined as the set of matrices which are upper triangular modulo N. Analogously we also define $\Gamma^0(N)$ as the group of all matrices lower triangular modulo N. A small computation shows that $\Gamma_0(4) \subset \sigma_2^{-1}\Gamma_\theta\sigma_2$, and therefore $\theta(2z)$ is a modular form for $\Gamma_0(4)$.

2.6. Bounds

The functional equation and the regularity at the cusps forces modular forms and their Fourier coefficients to have very particular growth rates. We give some basic results in this section. The proofs are based on the ones given in [**14**].

Assume f is a nonzero modular form of weight $k \geq 0$ for the group Γ. The non-negativity of the weight will be necessary.

PROPOSITION 2.8. *Let $\alpha_0 = k/2$ if f is cuspidal and $\alpha_0 = k$ otherwise. We have*

$$f(z) \ll (\Im z)^{-\alpha_0} \qquad as \quad \Im z \to 0^+.$$

Moreover this is sharp in the following sense:

(*i*) *For every irrational number x there is a constant $C_x > 0$ such that*

$$f(x+iy) \geq C_x\, y^{-k/2} \qquad \text{for infinitely many values of } y \to 0^+.$$

(*ii*) *For every rational x not cuspidal for f there is a constant $C_x > 0$ such that*

$$f(x+iy) \geq C_x\, y^{-k} \qquad \text{for infinitely many values of } y \to 0^+.$$

PROOF. We prove first the upper bound. Note it suffices to show the bound holds uniformly for $z = x + iy$ and $0 < y < 1/2$. Since the upper half-plane is covered by the Speiser circles $\mathcal{F}_{p/q}(2)$ (corollary 1.2), we can always find p/q such that $z \in \mathcal{F}_{p/q}(2)$. Applying the cusp asymptotics given by corollary 2.5 and the inequality $y|qz-p|^{-2} \geq 1/2$ provided by lemma 1.1 we obtain

$$f(z) \ll |f(p/q)|y^{-k} + y^{-k/2}.$$

If $x = p/q$ is a non-cuspidal rational point then the same expansion provided by corollary 2.5 shows $f(z) \gg y^{-k}$ as $y \to 0^+$, which is assertion (*ii*).

We show now (*i*). If x is an irrational number, by proposition 2.3 the vertical ray $\{\Re z = x\}$ intersects the boundary of infinitely many generalized Ford circles $\mathcal{F}_{p/q}(\delta)$ with $p/q \in [\infty]$ for δ big enough, in a sequence of points $z_n = x + iy_n$ where $y_n \to 0^+$. Now, the function $h(z) = (\Im z)^{k/2}|f(z)|$ is Γ-invariant, as can be readily checked from the functional equation (2.5) and (1.2), and therefore we must have $h(z_n) = h(z_n')$ for some z_n' lying in the boundary of $\mathcal{F}_\infty(\delta)$. This readily implies $|f(z_n)| \geq C y_n^{-k/2}$ for $C = \min_{\Im z = \delta^{-1}} h(z)$. To finish the proof it suffices to choose δ in such a way that f does not vanish on the line $\{\Im z = \delta^{-1}\}$, which is always possible as it is a periodic holomorphic function, guaranteeing $C > 0$. □

A sort of converse of this proposition is also true: the expansion at the cusps (theorem 2.4) shows that any modular function which is not a modular form grows at least exponentially fast when $\Im z \to 0^+$ on the Ford circles corresponding to some rationals. Some authors use this as a shortcut for defining modular forms; they can be defined as any modular function which grows at most polynomically fast when $\Im z \to 0^+$.

In order to derive bounds for the truncated Fourier series of a modular form we will use the fact the Dirichlet kernel satisfies the usual bounds even if evaluated in the complex plane but not too far away from the real line:

LEMMA 2.9. *Let $D_N(x) = \sum_{|n| \leq N} e(nx)$ and fix $y_0 > 0$. Denote by $\|\cdot\|_{\mathbb{Z}}$ the distance to the nearest integer. Then*

$$D_N(x+iy) \ll \min\left(N, \|x\|_{\mathbb{Z}}^{-1}\right),$$

uniformly for $x \in \mathbb{R}$ and $|y| \leq y_0/N$.

This can be shown as usual, by either trivially estimating the series or using the formula for the sum of a geometric series.

PROPOSITION 2.10. *Let $\alpha_0 = k/2$ if f is cuspidal and $\alpha_0 = k$ otherwise. The partial sums in the Fourier expansion given by theorem 2.4 satisfy*

$$\sum_{n \leq N} a_n e^{2\pi i (n+\kappa)x/m} \ll N^{\alpha_0} \log N,$$

uniformly for $x \in \mathbb{R}$.

PROOF. It suffices to prove the result for the expansion at ∞, as otherwise we may apply the result to $f|_\gamma$ instead, and composing with a scaling matrix we can moreover assume f is given by a Fourier series (2.8) with only integer frequencies, which converges uniformly on the sets $\Im z \geq \delta^{-1}$ for any $\delta \geq 0$. Hence

$$\sum_{n \leq N} a_n e^{2\pi i n x} = \int_0^1 f(u + i/N) D_N(x - u - i/N)\, du.$$

Applying the bounds obtained in proposition 2.8 and $\|D_N\|_1 \ll \log N$, which follows from lemma 2.9, we obtain the estimate. □

We provide in the next two propositions an estimation of the L^2-norm of the Fourier coefficients.

PROPOSITION 2.11. *If f is a cusp form the coefficients a_n of the Fourier expansion given by theorem 2.4 satisfy*

$$\sum_{n \leq N} |a_n|^2 \asymp N^k.$$

PROPOSITION. *If f is not a cusp form the coefficients a_n of the Fourier expansion given by theorem 2.4 satisfy*

$$\sum_{n \leq N} |a_n|^2 \asymp \phi(N)$$

where ϕ is given by

$$\phi(N) = \begin{cases} N^k & \text{if } 0 < k < 1, \\ N \log N & \text{if } k = 1, \\ N^{2k-1} & \text{if } k > 1. \end{cases}$$

We will only provide a proof for the cuspidal case, as we will not need the other result. For a self-contained proof of the non-cuspidal case we refer the reader to lemma 3.2 of [**14**]. Note that for weight $k > 1$ cusp forms always have coefficients which are much smaller than non-cuspidal forms, and in fact these can sometimes be interpretted as "error terms" in some problems in number theory. This is the case, for example, for generalizations of the 4-squares theorem (see §7.4 of [**82**]) or in the modularity theorem (see §4.4 of [**97**]). Note also that when applied to the discriminant function Δ this estimation can be interpreted as an average version of the Ramanujan conjecture $|\tau(p)| \leq 2p^{11/2}$.

PROOF OF PROPOSITION 2.11. Again it suffices to prove the result for the Fourier expansion at ∞ with only integer frequencies. We consider the Γ-invariant function $h(z) = (\Im z)^{k/2}|f(z)|$, which by virtue of proposition 2.10 and the exponential decay at ∞ it is uniformly bounded. Moreover we claim that we may find some constants C, $C' > 0$ such that $|\{x : h(x + i/N) > C\} \cap [0, 1]| > C'$ for every integer $N \geq 0$. Using Parseval's identity,

$$N^k \asymp N^k \int_0^1 |h(u + i/N)|^2\, du = \sum_{n \geq 0} |a_n|^2 e^{-4\pi n/N}.$$

The upper bound implies at once

$$\sum_{n\leq N}|a_N|^2 \ll N^k.$$

On the other hand, for any constant $K>0$, summing by parts and using the upper bound,

$$\sum_{n\geq KN}|a_n|^2e^{-4\pi n/N} \ll N^{k-1}e^{-2\pi K}\sum_{n\geq KN}(n/N)^k e^{-2\pi n/N} \ll N^k e^{-2\pi K},$$

and therefore for K big enough,

$$\sum_{n\leq KN}|a_n|^2 \geq \sum_{n\geq 0}|a_n|^2e^{-4\pi n/N} - \sum_{n>KN}|a_n|^2e^{-4\pi n/N} \gg N^k.$$

We still have to justify the previous claim. Let $0<C_1<C_2$ be constants to be determined later and consider the intervals $|x-p/q|\leq C_2/(qN^{1/2})$ with $C_1N^{1/2}<q<C_2N^{1/2}$. For $2C_2^3<C_1$ these are disjoint and cover a positive portion of the interval $[0,1]$. Suppose that $z=x+i/N$ with x lying in one of those intervals and let $\eta\in \mathrm{SL}_2(\mathbb{Z})$ satisfying $\eta(p/q)=\infty$. We may decompose $\eta^{-1}=\gamma\eta_i$, where $\gamma\in\Gamma$ and η_i lies in a fixed right-transversal for Γ. Hence $h(z)=h_i(\eta z)$ where $h_i(z)=h(\eta_i z)$. By (1.2) we have $1/(2C_2^2)\leq \Im(\eta z)\leq 1/C_1^2$ and therefore it suffices to show that we may choose C_1 and C_2 to ensure that every h_i is bounded below in that strip. This can be done by choosing $C_1\approx C_2$ both very small, as the Fourier expansion (2.8) of $f|_{\eta_i}(z)$ shows that this function cannot have any zeros when $\Im z$ is big enough. □

2.7. Bounds (II)

In this section we provide more specific bounds obtained by F. Chamizo and the author in [**20, 80**] with precise applications in mind. These will be crucial to obtain the results in chapters 3 and 5.

Our first result relates the growth of a modular form f near an irrational number with how close the rationals where f is not cuspidal are to the irrational in question.

PROPOSITION 2.12. *Let $\tau\geq 2$ and x_0 a fixed irrational number. The following holds:*

(*i*) *If all the rationals p/q not cuspidal for f satisfy*

$$\left|x_0-\frac{p}{q}\right| \gg \frac{1}{q^\tau} \tag{2.9}$$

then $f(x+iy)\ll y^{-\left(1-\frac{1}{\tau}\right)k}+y^{-k}|x-x_0|^{\frac{k}{\tau}}$ for $0<y<1/2$.

(*ii*) *If there are infinitely many rationals p/q not cuspidal for f satisfying*

$$\left|x_0-\frac{p}{q}\right| \ll \frac{1}{q^\tau} \tag{2.10}$$

then $f(x_0+iy)\gg y^{-\left(1-\frac{1}{\tau}\right)k}$ for infinitely many values of $y\to 0^+$.

PROOF. (*i*) Let $z=x+iy$ with $0<y<1/2$. Then z must be contained in one of the circles $\mathcal{F}_{p/q}(2)$. We will use again the expansion at the cusp given by corollary 2.5. If p/q is cuspidal for f then:

$$f(x+iy)\ll y^{-k/2}\leq y^{-\left(1-\frac{1}{\tau}\right)k}.$$

If p/q is not cuspidal we have

$$f(x+iy) \ll q^{-k}\left(\left(x-\frac{p}{q}\right)^2+y^2\right)^{-k/2}+y^{-\left(1-\frac{1}{\tau}\right)k}.$$

By hypothesis p/q satisfies (2.9) and therefore

$$q^{-k} \ll \left|x_0-\frac{p}{q}\right|^{k/\tau} \ll \left|x-\frac{p}{q}\right|^{k/\tau}+|x-x_0|^{k/\tau}.$$

Hence:

$$f(x+iy) \ll \left|x-\frac{p}{q}\right|^{k/\tau}\left(\left(x-\frac{p}{q}\right)^2+y^2\right)^{-k/2}+y^{-k}|x-x_0|^{k/\tau}+y^{-\left(1-\frac{1}{\tau}\right)k}.$$

Arguing by cases depending on whether $y \le |x-p/q|$ or not it readily shown that the first term is $\ll y^{-\left(1-\frac{1}{\tau}\right)k}$.

(ii) The case $\tau = 2$ has already been established in proposition 2.8, so we may assume $\tau > 2$. By hypothesis there must exist an equivalence class of non-cuspidal rationals modulo Γ for which infinitely many satisfy (2.10). For any of those rationals p/q we choose $z = x_0 + iy$ with $y = q^{-\tau}$ and note that

$$\frac{|qz-p|^2}{y} = q^{2+\tau}\left(\left|x_0-\frac{p}{q}\right|^2+y^2\right) \ll q^{2-\tau}.$$

Applying corollary 2.5 again we obtain:

$$|f(x_0+it)| = Cy^{-k/2}\left(\frac{y}{|qz-p|^2}\right)^{k/2}+O\left(y^{-k/2}e^{-Kq^{\tau-2}}\right) \gg y^{-k/2}q^{(\tau-2)k/2},$$

the constant $C = |f(p/q)|$ not depending on p/q. Using $q = y^{-1/\tau}$ the right hand side equals $y^{-\left(1-\frac{1}{\tau}\right)k}$. □

The other result we are going to include is a refinement of proposition 2.10 in the non-cuspidal case of weight 1 (although the proof can also be adapted to obtain refinements for other weights). It was inspired by the usual Hardy-Littlewood bound for a quadratic exponential sum:

$$\sum_{n=-N}^{N} e(n^2x) \ll \frac{N}{\sqrt{q}} \qquad \text{if} \qquad \left|2x-\frac{p}{q}\right| \le \frac{1}{qN} \quad \text{with} \quad q \le N. \tag{2.11}$$

A very simple proof with an extra error term $\sqrt{N}\log N$ can be consulted in §8.2 of [**62**]. The proof without the extra error term is much more demanding, and essentially follows from the original paper of Hardy and Littlewood on Diophantine approximation [**45**]. A more recent paper with an explicit statement and proof of this result is [**29**].

If we square the bound we obtain

$$\sum_{(n,m)\in Q} e((n^2+m^2)x) \ll \frac{N^2}{q} \qquad \text{with} \quad Q = [-N,N]\times[-N,N]. \tag{2.12}$$

We are going to need a similar bound but with the square Q replaced with a circle. Luckily, in that case, the series we are trying to bound is $\sum_{n\le N} r_2(n)e(nx)$, and we can exploit that this is a truncation of the Fourier series of θ^2, a modular form,

to give a very sharp estimate (in this regard, Hardy and Littlewood also exploit that θ is a modular form to obtain their bound in [**45**]). The idea is again to truncate the series by convolving by the Dirichlet kernel, integrating near the real line, as it was done in the proof of proposition 2.11. But this time the segment over which we are integrating will be broken into a Farey dissection, since by (1.11) in each subinterval we have a good approximation of the modular form by the cusp expansion. This is the principle behind the *circle method*, although when it is used to obtain asymptotics one can only estimate the integral well in an inner subinterval of each $\mathcal{A}_{p/q}$ (the so-called *major arcs*) and has to provide a rough upper bound in the remaining part (*minor arcs*). For our purposes we only need to consider one kind of arcs, greatly simplifying the proof.

PROPOSITION 2.13. *Let f be a modular form of weight 1, which admits an expansion as a Fourier series with only integer frequencies $f(z) = \sum_{n\geq 0} a_n e(nz)$. For every integer $N \geq 0$ we consider the Farey dissection of the continuum of order $\lfloor N^{1/2} \rfloor$. Then*

$$\sum_{n\leq N} a_n e^{2\pi i n x} \ll \frac{N(\log N)^2}{q + N|qx - p|} \qquad \text{if} \quad x \in \mathcal{A}_{p/q},$$

where the a_n are the Fourier coefficients of f and the O-constant only depends on f.

When applied to θ^2 this result shows that the bound (2.12) indeed holds when Q is replaced by a circle losing at most a power of a logarithm. Although the proof is not remarkably difficult, neither F. Chamizo nor I were able to find this result stated anywhere in the literature and was included in the article [**20**]. Surprisingly, shortly before the preprint was uploaded to arXiv a similar bound was uploaded in the preprint [**49**], but only applying to the case when sum is truncated by a smooth weight, which was not enough for our purposes.

For convenience we need some lemmas regarding the function

$$B(t) = \min\left(N, \|t\|_{\mathbb{Z}}^{-1}\right).$$

LEMMA 2.14. *With the same hypothesis as in proposition 2.13,*

$$f(x + i/N) \ll q^{-1} B(x - p/q) \qquad \text{if} \quad x \in \mathcal{A}_{p/q},$$

the $\ll$-constant only depending on f.

PROOF. By (1.11) the point $z = x + i/N$ lies inside the Speiser circle $\mathcal{F}_{p/q}(2)$. This means $\Im \gamma z \geq 1/2$ in the expansion provided by theorem 2.4, and the absolute convergence and the finiteness of the equivalence classes of cusps at once imply the uniform bound $|f(z)| \ll |j_\gamma(z)|^{-1} = q^{-1}|z - p/q|^{-1} \leq q^{-1} B(x - p/q)$. □

LEMMA 2.15. *For $t \in \mathbb{R}$ we have*

$$(B * B)(t) := \int_{-1/2}^{1/2} B(u) B(t-u)\, du \ll N \frac{\log(2 + N\|t\|_{\mathbb{Z}})}{2 + N\|t\|_{\mathbb{Z}}}.$$

PROOF. Cauchy's inequality gives $(B * B)(t) \ll \int_0^1 |B|^2 \ll N$. Using this and the symmetry, we can assume $2N^{-1} < t < 1/2$. If $0 < u < 1/2$ it is clear that the distance from t to u is smaller than the distance from t to $-u$. Hence $B(t-u) \geq B(t+u)$ and $(B*B)(t) \leq 2\int_0^{1/2} B(u)B(t-u)\, du$. This integral is less or equal than

$$\int_0^{N^{-1}} \frac{N\, du}{t-u} + \int_{N^{-1}}^{t-N^{-1}} \frac{du}{u(t-u)} + \int_{t-N^{-1}}^{t+N^{-1}} \frac{N\, du}{u} + \int_{t+N^{-1}}^{1/2+N^{-1}} \frac{du}{u(u-t)},$$

that gives $O(t^{-1}\log(Nt))$ evaluating or estimating the integrals. □

PROOF OF PROPOSITION 2.13. Assume for convenience $0 \le x < 1$. We have

$$\sum_{n\le N} a_n e(nx) = \int_0^1 f(u+i/N)D_N(x-u-i/N)\,du,$$

where D_N is the Dirichlet kernel. By lemmas 2.9 and 2.14

$$\sum_{n\le N} a_n e(nx) \ll \sum_{a/b} b^{-1} \int_{\mathcal{A}_{a/b}} B(u-a/b)B(x-u)\,du \tag{2.13}$$

where the sum ranges over the Farey sequence of $[0,1)$ of order $\lfloor N^{1/2}\rfloor$. Trivially

$$\mathcal{I}_{a/b} := \int_{\mathcal{A}_{a/b}} B(u-a/b)B(x-u)\,du \le (B*B)(x-a/b).$$

If $a/b = p/q$ we employ lemma 2.15 (with an extra logarithm to absorb an error term appearing later) to get

$$\mathcal{I}_{p/q} \ll \frac{N(\log N)^2}{1+N|x-p/q|}.$$

In the rest of the cases $\mathcal{I}_{a/b} \ll |x-a/b|^{-1}\log N$ also by lemma 2.15 (this is the best we can do as $|x-a/b| \gg N^{-1}$ by proposition 1.3). Substituting in (2.13)

$$\sum_{n\le N} a_n e(nx) \ll \frac{N(\log N)^2}{q+N|qx-p|} + (\log N)\sum_{a/b\neq p/q} |bx-a|^{-1}. \tag{2.14}$$

Each summand attains its maximum when x is one of the end-points of $\mathcal{A}_{p/q}$, both of which are rational numbers P/Q with $Q \asymp N^{1/2}$ (see proposition 1.3). Hence doubling the sum, it suffices to bound

$$\sum_{a/b\neq p/q} |bP/Q-a|^{-1} = Q\sum_{m\le 2N} m^{-1}\#\{a/b \,:\, Pb-Qa = \pm m\}.$$

The last cardinality is $O(1)$ because given any two solutions of $Pb_i - Qa_i = m$ (or $-m$) the difference b_1-b_2 is a multiple of Q, but $b_i \le N^{1/2}$. Introducing this bound in (2.14), the result follows. □

In fact, since we will need the bound we have just proved for functions which are not modular forms, we state for convenience the facts we have used for the proof:

COROLLARY 2.16. *Proposition 2.13 is true for any function having a Fourier expansion uniformly converging on the sets $\{\Im z \ge \delta^{-1}\}$ and satisfying the bound in lemma 2.14.*

2.8. Theta functions

Let Q be an integral binary quadratic form (**I**.14) which is positive definite. For every integer $n \ge 0$ let $r_Q(n)$ denote the number of representations of n by Q. We are going to show that the function

$$\theta_Q(z) = \sum_{\vec{n}\in\mathbb{Z}^2} e^{2\pi i Q(\vec{n})z} = \sum_{n\ge 0} r_Q(n)e(nz)$$

is a modular form of weight 1 for some congruence group and some multiplier system. Note in the particular case $Q(x,y) = x^2+y^2$ this function coincides with $\theta^2(2z)$, which we have already seen to be a modular form of weight 1 for $\Gamma_0(4)$.

This result is usually presented in a more general form, as $\sum_{\vec{n}} P(\vec{n})e(Q(\vec{n})z)$ is a modular form of weight $r/2 + \deg P$ whenever P is an homogeneous polynomial harmonic with respect to Q and Q a positive definite integral quadratic form on r variables (see §10 of [**61**]). Here however, we need a generalization in another direction. Let $\vec{v} = (\alpha, \beta) \in \mathbb{R}^2$ and consider

$$r_{Q,\vec{v}}(n) = \sum_{Q(n_1,n_2)=n} e(\alpha n_1 + \beta n_2). \tag{2.15}$$

The function

$$\theta_{Q,\vec{v}}(z) = \sum_{\vec{n}\in\mathbb{Z}^2} e^{2\pi i Q(\vec{n})z + 2\pi i \vec{n}\cdot\vec{v}} = \sum_{n\geq 0} r_{Q,\vec{v}}(n)e(nz) \tag{2.16}$$

is holomorphic in the upper half-plane and transforms in a very similar way to a theta function. In fact for some special values of $\vec{v}$ it coincides with some of the so-called *Jacobi modular forms*, and in particular $\theta_{Q,\vec{0}} = \theta_Q$. We are going to derive the general transformation formula, adapted from chapter 4 of Siegel's notes [**87**].

Some notation first. There is a unique symmetric matrix A with integer coefficients such that $Q(\vec{x}) = \frac{1}{2}\vec{x}^t A\vec{x}$. The inverse matrix A^{-1}, however, need not have integer coefficients, but its entries are rationals of denominator dividing $\det A \geq 1$. In particular, the quotient $A^{-1}\mathbb{Z}^2/\mathbb{Z}^2$ is well-defined and finite, and in fact contains exactly $\det A$ elements. Let $\mathcal{L}$ be a set of representatives of this quotient; one such set can be constructed by taking all the elements in Λ lying in the square $[0,1)\times[0,1)$. For every member $\vec{\ell} \in \mathcal{L}$ and $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$ not fixing ∞ we define the Gauss sum

$$G_\gamma(\vec{\ell}) = \frac{1}{c} \sum_{\vec{g} \pmod c} e\left(-\frac{aQ(\vec{\ell}+\vec{g})}{c}\right),$$

where the sum runs over a complete set of representatives of $\mathbb{Z}^2/c\mathbb{Z}^2$. Indeed, expanding Q by the formula $Q(\vec{x}+\vec{y}) = Q(\vec{x}) + \vec{x}^t A\vec{y} + Q(\vec{y})$ and using that $A\vec{\ell}$ has integer components it is clear that each summand does not depend on the choice of the representative $\vec{g}$, and therefore changing the representative of $\vec{\ell}$ amounts to reordering the sum.

THEOREM 2.17. *Let $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ not fixing ∞ and $\vec{u} = A^{-1}\vec{v}$. Then*

$$j_\gamma(z)\theta_{Q,\vec{v}}(z) = \frac{\delta(\gamma,\vec{v})}{\sqrt{\det A}} \sum_{\vec{\ell}\in\mathcal{L}} G_\gamma(\vec{\ell}) \sum_{\vec{x}\in\mathbb{Z}^2+\vec{\ell}} e(Q(\vec{x}+c\vec{u})\gamma z - a\vec{x}\cdot\vec{v}),$$

where $\delta(\gamma,\vec{v})$ is a unimodular constant.

We will prove this theorem at the end of the section. The Gauss sums involved satisfy the following properties

LEMMA 2.18. *We have $|G_\gamma(\vec{\ell})| \leq \sqrt{\det A}$. Moreover if $2N \mid c$ where N is any positive integer satisfying that the matrix NA^{-1} has integer entries, then $|G_\gamma(\vec{0})| = \sqrt{\det A}$ and $G_\gamma(\vec{\ell}) = 0$ for $\vec{\ell} \notin \mathbb{Z}^2$.*

PROOF. The proof of the upper bound mimics the classical one. Squaring the Gaussian sum,

$$|G_\gamma(\vec{\ell})|^2 = \frac{1}{c^2} \sum_{\vec{g}_1,\vec{g}_2 \pmod c} e\left(\frac{a\vec{\ell}^t A(\vec{g}_2-\vec{g}_1) + aQ(\vec{g}_2) - aQ(\vec{g}_1)}{c}\right).$$

Writing $\vec{h} = \vec{g}_2 - \vec{g}_1$,

$$|G_\gamma(\vec{\ell})|^2 = \frac{1}{c^2} \sum_{\vec{h} \pmod c} e\left(\frac{a\vec{\ell}^t A\vec{h} + aQ(\vec{h})}{c}\right) \sum_{\vec{g}_1 \pmod c} e\left(\frac{a\vec{h}^t A\vec{g}_1}{c}\right).$$

Let $\vec{w} = aA\vec{h}$, and note that the sum $\sum_{\vec{g}_1} e(\vec{w}^t\vec{g}_1/c)$ vanishes unless $\vec{w} \equiv \vec{0} \pmod c$. Hence

$$|G_\gamma(\vec{\ell})|^2 = \sum_{\substack{\vec{h} \pmod c \\ A\vec{h} \equiv \vec{0} \pmod c}} e\left(\frac{a\vec{\ell}^t A\vec{h} + aQ(\vec{h})}{c}\right).$$

We are summing over those $\vec{h} \in \mathbb{Z}^2$ modulo $c\mathbb{Z}^2$ satisfying $A\vec{h} \in c\mathbb{Z}^2$, hence on a subset of representatives of $cA^{-1}\mathbb{Z}^2/c\mathbb{Z}^2$. Hence the sum has at most $\#\mathcal{L} = \det A$ summands.

When $2N \mid c$ then on the one hand all the members of $cA^{-1}\mathbb{Z}^2$ have integer coordinates, and therefore the sum has exactly $\det A$ summands; and on the other $\vec{h} = cA^{-1}\vec{h}_1$ implies $c \mid Q(\vec{h})$. Now, the remaining sum can be seen as a character of the group $cA^{-1}\mathbb{Z}^2/c\mathbb{Z}^2$ summed over the whole group, and hence vanishes if and only if for some $\vec{h} \in cA^{-1}\mathbb{Z}^2$ we have $e(a\vec{\ell}^t A\vec{h}/c) \neq 1$. If $\vec{\ell} \notin \mathbb{Z}^2$ then we can assume that one of its components must lie in the strict interval $(0,1)$, say the first, and then we can choose $\vec{h} = cA^{-1}\vec{e}_1$ for $\vec{e}_1^t = (1,0)$. Note $e(a\ell_1) \neq 1$ as a is coprime to c and the denominator of ℓ_1 must be a divisor of $2N$, and therefore of c. □

Combining lemma 2.18 and theorem 2.17 the following two corollaries are immediate:

Corollary 2.19. *Let N be a positive integer such that NA^{-1} has integer entries. Then θ_Q is a modular form of weight 1 for $\Gamma_0(2N)$.*

Corollary 2.20. *The truncation of the Fourier series defining $\theta_{Q,\vec{v}}$ satisfies the bounds of proposition 2.13 uniformly in $\vec{v} \in \mathbb{R}^2$.*

The first corollary follows from the transformation law, which in this case reads $j_\gamma(z)\theta_Q(z) = \mu_\gamma^{-1}\theta_Q(\gamma z)$ when $\gamma \in \Gamma_0(2N)$. The same formula also shows that $\lim_{\Im z \to \infty} \theta_Q|_\gamma(z) = i(\det A)^{-1/2} G_{\gamma^{-1}}(\vec{0})$ for any $\gamma \in \mathrm{SL}_2(\mathbb{Z})$. Note that in deriving these results we have not used anything essential of binary forms, and indeed the proof may be easily adapted to cover the case of theta functions associated to n-ary quadratic forms.

The second corollary follows from estimating the exponential sum in the transformation law termwise to show that the right hand side is uniformly bounded and therefore we can apply corollary 2.16.

The proof of the transformation law is essentially Poisson summation, in the form of the following lemma, applied to the sum defining $\theta_{Q,\vec{v}}$ in arithmetic progressions.

Lemma 2.21. *For any $z \in \mathbb{Z}$ and $\vec{x} \in \mathbb{C}^2$ we have*

$$\sum_{\vec{n} \in \mathbb{Z}^2} e(Q(\vec{n} + \vec{x})z) = \frac{i}{z\sqrt{\det A}} \sum_{\vec{n} \in \mathbb{Z}^2} e(-Q(A^{-1}\vec{n})/z + \vec{n} \cdot \vec{x})$$

Proof. By the identity principle it suffices to show the result holds for $z = it$. We are therefore going to apply Poisson summation in two variables to the function $f(\vec{u}) = \exp\{-2\pi Q(\vec{u} + \vec{x})t\}$. Note that since A is real, symmetric and positive

definite there exists a nonsingular matrix L such that $A = L^t L$. Therefore, if we let $h(\vec{u}) = \exp\{-\pi \vec{u}\cdot\vec{u}\}$ then $f(\vec{u}) = h(L(\vec{u}+\vec{x})\sqrt{t})$. Using $\hat{h} = h$ and elementary properties of the Fourier transform we can compute

$$\hat{f}(\vec{\xi}) = \frac{e(\vec{x}\cdot\vec{\xi})}{t\sqrt{\det A}} e^{-2\pi Q(A^{-1}\vec{\xi})/t}.$$

The identity $\sum_{\vec{n}} f(\vec{n}) = \sum_{\vec{n}} \hat{f}(\vec{n})$ is precisely the one stated for $z = it$. □

Proof of theorem 2.17. By the definition of $\theta_{\vec{v}}(z)$ and separating classes modulo c,

$$\theta_{Q,\vec{v}}(z) = \sum_{\vec{g} \pmod c} \sum_{\vec{m}\in\mathbb{Z}^2} e\big(Q(c\vec{m}+\vec{g})z + (c\vec{m}+\vec{g})\cdot A\vec{u}\big).$$

Writing $(j_\gamma(z)-d)/c$ instead of z and completing squares, the phase can be expressed as $P_1 + P_2$ with

$$P_1 = \frac{j_\gamma(z)}{c} Q\left(c\vec{m}+\vec{g}+\frac{c\vec{u}}{j_\gamma(z)}\right) \qquad \text{and} \qquad P_2 = -\frac{c}{j_\gamma(z)}Q(\vec{u}) - \frac{d}{c}Q(c\vec{m}+\vec{g}).$$

Note that P_2 does not change modulo 1 when $\vec{m}$ varies and we can put $\vec{m} = \vec{0}$. On the other hand, by lemma 2.21,

$$\sum_{\vec{m}\in\mathbb{Z}^2} e(P_1) = \frac{i(\det A)^{-1/2}}{c j_\gamma(z)} \sum_{\vec{m}\in\mathbb{Z}^2} e\left(-\frac{Q(A^{-1}\vec{m})}{c j_\gamma(z)} + c^{-1}\Big(\vec{g}+\frac{c\vec{u}}{j_\gamma(z)}\Big)\cdot\vec{m}\right).$$

Under the change of variables $\vec{x} = A^{-1}(-\vec{m})$ with $\vec{x} = \vec{n} + \vec{\ell}$, where $\vec{\ell}\in\mathcal{L}$ and $\vec{n}\in\mathbb{Z}^2$, this phase corresponds to

$$P_3 = -\frac{Q(\vec{x}) + c\vec{v}\cdot\vec{x}}{c j_\gamma(z)} - c^{-1}\vec{g}^t A\vec{x}.$$

Let $w = \gamma z$. Substituting $\big(j_\gamma(z)\big)^{-1} = j_{\gamma^{-1}}(w) = -cw + a$ in P_2 and P_3,

$$e(P_2+P_3) = e\Big(wQ(\vec{x}) + (cw-a)\vec{v}\cdot\vec{x} + c(cw-a)Q(\vec{u})\Big)\, e\Big(-\frac{a}{c}Q(\vec{x}) - \frac{1}{c}\vec{g}^t A\vec{x} - \frac{d}{c}Q(\vec{g})\Big).$$

The last exponential is $e(-aQ(\vec{x}+d\vec{g})/c)$ because $ad \equiv 1 \pmod c$ and $A\vec{x}$ has integer coefficients, and when we sum on $\vec{g}$ we obtain $cG_\gamma(\vec{\ell})$. It only remains to note that the argument of the first exponential can be written as $wQ(\vec{x}+c\vec{u}) - a\vec{v}\cdot\vec{x} - acQ(\vec{u})$. □

2.9. Hecke newforms

In this section we provide a glimpse of the Atkin-Lehner theory of Hecke newforms. These objects will appear as examples in chapter 3.

In §2.5 we saw that if f is a modular form for $\Gamma_0(N)$ then both f and $f|_{\sigma_m}$ are modular forms for $\Gamma_0(mN)$. Hence when we study the space of forms of a given weight for $\Gamma_0(N)$ and a given multiplier system there might be some forms which are not really new, but come from forms in $\Gamma_0(d)$ for some divisor $d \mid N$. We would like to "remove" these forms and the subspace generated by them, but of course in general the complement of a vector subspace is not unique. When we restrict our attention to cusp forms, however, there is a well-defined inner product in the space of modular forms: $\langle f, g\rangle = \int f\bar{g}\,(\Im z)^k\, d\mu$ where μ is the hyperbolic measure and we integrate over a fundamental domain for $\Gamma_0(N)$ (see §5.2 of [**82**]). We can therefore take orthogonal complements with respect to this inner product.

Moreover, for some particular choices of "nice" multiplier systems, there is a rich theory of arithmetic operators acting on the space of cusp forms, called the *Hecke operators*. We are not going to define them here, we refer the reader instead to chapter 9 of [**82**]. In particular when the weight is an even integer and $\mu_\gamma = \chi(d)$ where χ is a Dirichlet character modulo N and d the lowest-right entry of γ (modular forms of *nebentypus* χ), the Hecke operators are normal operators with respect to the inner product and commute with each other, and therefore any invariant subspace admits a basis of eigenvectors of all the Hecke operators. In particular cusp forms which are eigenvectors of all Hecke operators are referred to as *eigenforms*.

Let M be space of cusp forms for $\Gamma_0(N)$ and nebentypus χ, and let M^- be the space spanned by all cusp forms arising from groups $\Gamma_0(d)$ for $d \mid N$, $d < N$. Then $M = M^- \oplus M^+$ where M^+ is the orthogonal complement to M^-. Both spaces are invariant under the action of the Hecke operators, and therefore admit a basis of eigenforms. For M^+ these eigenform can be seen to be uniquely determined up to constant, and therefore we can take a canonical representative whose first nonzero Fourier coefficient in the expansion at ∞ is normalized to be 1. Each of these special eigenforms are called *newforms* and provide a canonical basis for M^+. When acted upon by some $|_{\sigma_d}$ for $d > 1$ these belong to the space M^- of a smaller group, and they are referred to as *oldforms*. The oldforms can be seen to generate M^-, and hence together with the newforms provide a canonical basis for the whole space of cusp forms.

Newforms are very important objects in modern algebraic number theory. The online database [**76**] can be used to explore them and their relation to other mathematical objects, specially elliptic curves, for low level groups.

CHAPTER 3

Regularity of fractional integrals of modular forms

The contents of this chapter comprise the results of the article "On the regularity of fractional integrals of modular forms" [**80**], and will be presented more or less in the same order. The article represents the continuation of the research line started by F. Chamizo in [**14**] and continued by Chamizo, Petrykiewicz and Ruiz-Cabello in [**19**] and by Ruiz-Cabello in [**83**].

3.1. Hölder exponents

The regularity of a function may be studied in many different ways depending on the applications in mind. In our case we are going to choose the same notions that were already considered by Chamizo, Petrykiewicz and Ruiz-Cabello in [**19**]. These are inspired by the work of Jaffard, Seuret and Véhel in multifractal analysis [**65, 86**].

The regularity will be measured in terms of different *Hölder exponents*, but in order to define them first we need to introduce some function spaces. The functions considered will be complex valued, defined in either all $\mathbb{R}$ or in an open subset of $\mathbb{R}$.

- For $0 \leq s \leq 1$ we define $\Lambda^s(x_0)$ as the set of all continuous functions which satisfy a s-Hölder condition at x_0, *i.e*,

$$|f(x) - f(x_0)| \ll |x - x_0|^s \qquad \text{as} \quad x \to x_0.$$

  We analogously define $\Lambda^s(\Omega)$ for a subset $\Omega \subset \mathbb{R}$ as the set of all continuous functions satisfying a uniform s-Hölder condition on Ω.

- For any $s \geq 0$ we define $\mathcal{C}^s(x_0)$ as the set of all continuous functions for which there is some polynomial P satisfying

$$|f(x) - P(x - x_0)| \ll |x - x_0|^s \qquad \text{as} \quad x \to x_0.$$

  Note that we can always assume P is of degree smaller than s.

- For any $0 \leq s \leq 1$ and any integer $k \geq 0$ we define $\mathcal{C}^{k,s}(x_0)$ as the set of all continuous functions for which $f^{(k)}$ exists in an open interval I containing x_0 and verifies $f^{(k)} \in \Lambda^s(x_0)$. Analogously one defines $\mathcal{C}^{k,s}(\Omega)$ for an open set $\Omega \subset \mathbb{R}$ as the set of all continuous functions for which $f^{(k)}$ exists in Ω and $f^{(k)} \in \Lambda^s(K)$ for every compact subset $K \subset \Omega$.

Finally we also define the spaces $\Lambda^s_{\log}$, $\mathcal{C}^s_{\log}$ and $\mathcal{C}^{k,s}_{\log}$ by replacing $|x - x_0|^s$ in the previous definitions with $|x - x_0|^s \log |x - x_0|$.

Note for $0 \leq s \leq 1$ we have $\Lambda^s(x_0) = \mathcal{C}^s(x_0) = \mathcal{C}^{0,s}(x_0)$, and hence these spaces constitute different generalizations of the notion of Hölder continuity. The three Hölder exponents β, β^* and β^{**} are then defined in the following way:

$$\beta(x_0) := \sup\{s \,:\, f \in \mathcal{C}^s(x_0)\},$$
$$\beta^*(x_0) := \sup\{k + s \,:\, f \in \mathcal{C}^{k,s}(x_0)\},$$

$$\beta^{**}(x_0) := \lim_{I \to \{x_0\}} \sup\{k+s \,:\, f \in \mathcal{C}^{k,s}(I)\}.$$

In the last definition the limit is taken as I runs over a sequence of nested open intervals whose intersection is $\{x_0\}$.

The first exponent, $\beta(x_0)$, also called the *pointwise Hölder exponent*, is the most local in nature and gives precise information about how well the function can be approximated by a polynomial in arbitrarily small neighborhoods of x_0, even when no derivative exists near that point (note that P in the definition of $\mathcal{C}^s(x_0)$ generalizes the notion of Taylor polynomial when f cannot be differentiated $\lceil s \rceil - 1$ times). This exponent also has the advantage of being the most easily studied through the tool of the wavelet transform [**65**], as we will see in §3.4.

The second one, $\beta^*(x_0)$, also called the *restricted local Hölder exponent*, is more demanding in the sense that f must be differentiable enough times for it to coincide with $\beta(x_0)$. This is in some sense like imposing that the polynomial is the usual Taylor polynomial of f. It was introduced in [**19**] as a compromise between the exponents β and β^{**}.

Finally, $\beta^{**}(x_0)$, the *local Hölder exponent*, requires f not only to be differentiable in open neighborhoods, but also its k-th derivative to satisfy a Hölder condition in them. The importance of this last one resides in the fact that it behaves well under the action of a wide class of pseudo-differential operators, and for this reason it was introduced by Seuret and Véhel in [**86**].

The inclusions $\mathcal{C}^{k,s}(I) \subset \mathcal{C}^{k,s}(x_0) \subset \mathcal{C}^{k+s}(x_0)$, the last one a consequence of Taylor's theorem, imply that these exponents satisfy the inequalities

$$\beta(x) \geq \beta^*(x) \geq \beta^{**}(x).$$

These inequalities are, in general, strict. For example the function f defined by $f(x) = x^4 \sin(x^{-2})$ if $x \neq 0$ and $f(0) = 0$ has $\beta(0) = 4 > \beta^*(0) = 2 > \beta^{**}(0) = 4/3$.[1] We can even have $\beta(x_0) = \infty$ and $\beta^{**}(x_0) = 0$ for more extreme examples such as $f(x) = e^{-x^2} \sin(e^{x^{-4}})$ and $f(0) = 0 = x_0$.

3.2. Main results

Let f be a nonzero modular form of weight[2] $r > 0$ for a finite index subgroup Γ of $\mathrm{SL}_2(\mathbb{Z})$ and multiplier system μ. Then f has a Fourier expansion at ∞ (*cf.* (2.8))

$$f(mz) = \sum_{n \geq 0} a_n e^{2\pi i (n+\kappa) z}. \tag{3.1}$$

Given $\alpha > 0$ we define the α-fractional integral of f as the formal series (*cf.* [**14, 19, 24, 64, 83**])

$$f_\alpha(mx) := \sum_{n+\kappa > 0} \frac{a_n}{(n+\kappa)^\alpha} e^{2\pi i (n+\kappa) x}. \tag{3.2}$$

[1] The only exponent difficult to compute is $\beta^{**}(0)$. To see it equals 4/3, note that for $x \neq 0$ we have $f'(x) = 4x^3 \sin(x^{-2}) - 2x \cos(x^{-2})$ and taking $x_n^{-2} = \pi n$ and $y_n^{-2} = \pi(n+1)$ we see that $\sup_n |f'(x_n) - f'(y_n)|/|x_n - y_n|^s = \infty$ for any $s > 1/3$. On the other hand, $f''(x) = O(x^{-2})$ and by the mean value theorem $0 \leq x \leq y$ implies $|f'(x) - f'(y)|/|x-y|^{1/3} \ll x^{-2}|x-y|^{2/3}$. This is bounded if $|x-y| \leq x^3$. Otherwise use $f'(x) = O(x)$ to bound the incremental quotient by $y/|x-y|^{1/3}$. Either $y \leq 2x$ or $y - x \geq y/2$, and hence this latter expression is also bounded.

[2] The letter k is traditionally reserved for the weight of the modular form. In this chapter, however, we will use r to avoid the notational clash with the functional spaces defined above.

For example, $\Im\theta_1(x) = 2\varphi(x)$ where θ is Jacobi's theta function (**I**.1) and φ is Riemann's example (**I**.9).

For any $\gamma \in \mathrm{GL}_2^+(\mathbb{R})$ such that $\gamma^{-1}\Gamma\gamma \cap \mathrm{SL}_2(\mathbb{Z})$ has finite index in $\mathrm{SL}_2(\mathbb{Z})$ the function $f|_\gamma$ is again a modular form and we can also form $(f|_\gamma)_\alpha$. To avoid excessive use of subscripts we are going to introduce the nonstandard notation f^γ to mean the same as $f|_\gamma$, and then we are going to define $f_\alpha^\gamma := \left(f^\gamma\right)_\alpha$. In particular we may always choose $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ with $\gamma\infty = \mathfrak{a}$ for any cusp $\mathfrak{a} \in \mathbb{Q} \cup \{\infty\}$, producing a collection of related formal series

$$f_\alpha^\gamma(m_\mathfrak{a} x) = \sum_{n+\kappa_\mathfrak{a}>0} \frac{a_n^\mathfrak{a}}{(n+\kappa_\mathfrak{a})^\alpha} e^{2\pi i(n+\kappa_\mathfrak{a})x}.$$

From the remarks in §2.4 it follows that f_α^γ is uniquely determined by the orbit of the cusp $\mathfrak{a}$ modulo Γ up to translation and multiplication by an unimodular constant.

Our first three theorems establish some global and local regularity properties of f_α. We define throughout this chapter $\alpha_0 := r/2$ if f is a cusp form and $\alpha_0 := r$ otherwise.

Theorem 3.1 (Global regularity). *Let $\alpha > 0$. The following holds:*

(*i*) *If $\alpha \le \alpha_0$ the series* (3.2) *defining f_α diverges in a dense set.*

(*ii*) *If $\alpha > \alpha_0$ the series* (3.2) *defining f_α converges uniformly to a continuous function in all the real line. Moreover $f_\alpha \in \mathcal{C}^{\lfloor\alpha-\alpha_0\rfloor,\{\alpha-\alpha_0\}}(\mathbb{R})$ if $\alpha-\alpha_0 \notin \mathbb{Z}$ and $f_\alpha \in \mathcal{C}_{\log}^{\alpha-\alpha_0-1,1}(\mathbb{R})$ otherwise.*

(*iii*) *If $0 < \alpha - \alpha_0 \le 1$ then $f_\alpha \notin \mathcal{C}^{1,0}(I)$ for any open interval I. The same is true for $\Re f_\alpha$ and $\Im f_\alpha$.*

The statements in this theorem concerning the convergence or divergence of the series (3.2) were known (proposition 3.1 of [**14**]). Part three is a generalization of lemma 3.5 of [**19**].

For the remaining results stated in this section we will assume $\alpha > \alpha_0$.

Theorem 3.2 (Local regularity at rationals). *Let x be a rational number and $\beta(x), \beta^*(x)$ and $\beta^{**}(x)$ the Hölder exponents of either f_α, $\Re f_\alpha$ or $\Im f_\alpha$. Then:*

(*i*) *If f is cuspidal at x then $\beta(x) = 2\alpha - r$. Otherwise $\beta(x) = \alpha - r$.*

(*ii*) *If f is a cusp form then*

$$\beta^*(x) = [\alpha - r/2] + \min\left(1, 2\{\alpha - r/2\}\right).$$

If f is not a cusp form then

$$\beta^*(x) = \begin{cases} \lfloor\alpha - r\rfloor + \min\left(1, 2\{\alpha - r\} + r\right) & \text{if } f \text{ cuspidal at } x \text{ and } \alpha - r \notin \mathbb{Z}, \\ \alpha - r & \text{if } f \text{ not cuspidal at } x \text{ or } \alpha - r \in \mathbb{Z}. \end{cases}$$

(*iii*) *In any case $\beta^{**}(x) = \alpha - \alpha_0$.*

(*iv*) *If $0 < \alpha - \alpha_0 \le 1$ then f_α (resp. $\Re f_\alpha$, $\Im f_\alpha$) is not differentiable at any rational point which is not cuspidal for f. If x is cuspidal for f then f_α (resp. $\Re f_\alpha$, $\Im f_\alpha$) is differentiable at x if and only if $\alpha > (r+1)/2$, and in this case the derivative is given by*

$$f'_\alpha(x) = \frac{(2\pi)^\alpha}{(im)^\alpha \Gamma(\alpha)} \int_{(x)} (z-x)^{\alpha-1} f'(z)\, dz,$$

where (x) denotes the vertical ray connecting x with $i\infty$, and the symbol $\Gamma(\cdot)$ stands for the gamma function (not to be confused with the group Γ associated to f).

Our previous knowledge on these Hölder exponents at rational points was very poor, specially in the non-cuspidal case (*cf.* theorems 3.3, 3.4 and 3.6 of [**19**]). Part (*iv*) is essentially contained in theorem 2.2 of [**14**].

The regularity at irrational points depends on how well these points can be approximated by rationals which are not cuspidal for f. This is precisely measured by the following quantity:[3]

$$\tau_x := \sup\left\{\tau \,:\, \left|x-\frac{p}{q}\right| \ll \frac{1}{q^\tau} \text{ for infinitely many non-cuspidal rationals } \frac{p}{q}\right\}. \tag{3.3}$$

Note that the inequality $\tau_x \geq 2$ is always satisfied for any irrational number x and, in fact, the number 2 is always contained in the set on the right hand side of (3.3), as shown by proposition 2.3. On the other hand, when $\tau_x = \infty$ we establish the convention $1/\tau_x = 0$.

THEOREM 3.3 (LOCAL REGULARITY AT IRRATIONALS). *Let x be any irrational number and $\beta(x), \beta^*(x)$ and $\beta^{**}(x)$ the Hölder exponents of either f_α, $\Re f_\alpha$ or $\Im f_\alpha$. Then:*

(*i*) *If f is a cusp form then $\beta(x) = \beta^*(x) = \beta^{**}(x) = \alpha - r/2$.*

(*ii*) *If f is not a cusp form,*

$$\begin{aligned}
\beta(x) &= \alpha - \left(1-\frac{1}{\tau_x}\right) r, \\
\beta^*(x) &= \begin{cases} \lfloor \alpha - r \rfloor + \min\left(1, \{\alpha - r\} + r/\tau_x\right) & \text{if } \alpha - r \notin \mathbb{Z}, \\ \alpha - r & \text{if } \alpha - r \in \mathbb{Z}, \end{cases} \\
\beta^{**}(x) &= \alpha - r.
\end{aligned}$$

REMARK. *Regarding the differentiability of these functions at irrational points we could not prove anything beside the obvious results: it cannot be differentiable whenever $\beta(x) < 1$, while it must be for $\beta(x) > 1$.*

The cuspidal case was already covered by theorem 3.1 of [**19**], while the non-cuspidal case was previously only known for "Riemann's example" [**64**] and for modular forms for $\Gamma_0(N)$ under strong restrictions on α, result contained in propositions 3.15 and 3.17 of [**83**].

We have defined f_α^σ as $(f^\sigma)_\alpha$, and although these operators do not commute the function $(f^\sigma)_\alpha$ is closely related to $(f_\alpha)^\sigma$. The relation takes the shape of an approximate functional equation for f_α, resembling the one for f but modulo a reasonably good error term. This approximate functional equation not only plays a key role in the proof of the theorems stated above, but also has interest on its own.

THEOREM 3.4 (APPROXIMATE FUNCTIONAL EQUATION). *Let $\sigma \in \mathrm{SL}_2(\mathbb{R})$ satisfying that f^σ is a modular form and $x_0 = \sigma\infty \in \mathbb{Q}$. Assume moreover that the lowest-left*

[3]The symbol $\ll$ could be replaced by $\leq$ in this definition without affecting the value of τ_x, but this convention simplifies some arguments later on.

*entry of σ is negative (*i.e. σ^{-1} *satisfying* (2.6)*). Then there exist two nonzero real constants A, B with $B > 0$, depending on σ, such that:*

$$f_\alpha(x) = Ai^{-\alpha} f(x_0)\phi(x - x_0) + B|x - x_0|^{2\alpha}(x - x_0)^{-r} f_\alpha^\sigma(\sigma^{-1}x) + E(x)$$

where $f(x_0) = \lim_{\Im z\to\infty} f^\sigma(z)$ and

$$\phi(x) = \begin{cases} x^{\alpha-r} & \text{if } \alpha - r \notin \mathbb{Z}, \\ x^{\alpha-r}\log x & \text{if } \alpha - r \in \mathbb{Z}. \end{cases}$$

The error term $E(x)$ lies in the spaces $\mathcal{C}^{1,0}(\mathbb{R} \setminus \{x_0\})$ and $\mathcal{C}^{2\alpha-r+1}(x_0)$.

When $\sigma \notin \mathrm{SL}_2(\mathbb{Z})$ the value of $f(x_0)$ considered in this theorem might not coincide with the one we have defined in §2.4; they differ in the nonzero complex constant $\lim_{\Im z\to\infty} \left(j_\gamma(\gamma^{-1}\sigma z)\right)^r \left(j_\sigma(z)\right)^{-r}$.

The error term in the approximate functional equation is essentially a polynomial close to the point x_0, and hence as $x \to x_0$ the graph of f_α looks like a deformed version of that of f_α^σ. As the latter is a periodic function and $\sigma x \to \infty$, this gives these functions a "fractal look", where some motif gets repeated an infinitude of times near every rational (*cf.* figure **I**.2). When the theorem is applied to f^γ and $\sigma = \gamma^{-1}\beta$ for some $\gamma, \beta \in \mathrm{SL}_2(\mathbb{Z})$ it also relates the graphs of f_α^γ (close to the rational $x_0 = \sigma\infty$) and f_α^β (globally). This will be explored in more detail in §3.7. When $\sigma \in \Gamma$ we have $f^\sigma = \mu_\sigma f$, and if $f(x_0) = 0$ the approximate functional equation looks almost like (2.5). In this particular case the theorem essentially corresponds to lemma 3.8 of [**19**], while for "Riemann's example" φ it was originally obtained by Duistermaat in [**24**].

Another particular case of Theorem 3.4 was known in the literature: when f is a classical cusp form of even integer weight $r > 2$ and $\alpha = r - 1$ the function f_{r-1} is known as the Eichler integral of f and the approximate equation is in fact exact, the error term corresponding to the period polynomial of f of the Eichler-Shimura theory (*cf.* [**28**]). We are going to recover this result as a corollary of (the proof) of theorem 3.4:

Corollary 3.5. *If f is a cusp form of weight $r > 2$ and $\alpha = r - 1$ then the error term $E(x)$ in theorem 3.4 is given by*

$$E(x) = \frac{(2\pi)^{r-1}}{(im)^{r-1}\Gamma(r-1)} \int_{(x_0)} (z - x)^{r-2} f(z)\, dz.$$

If moreover r is an integer then E is a polynomial.

Theorem 3.3 shows that when f is not a cusp form the pointwise Hölder exponent β of f_α at the irrational numbers ranges in a continuum between the values $\alpha - r$ and $\alpha - r/2$. An interesting concept to study in this case is that of the spectrum of singularities, which measures in some rough sense how big are the sets of points where each Hölder exponent is attained. It is defined as the function $d : [0, +\infty) \to [0, 1] \cup \{-\infty\}$ associating to each $\delta \geq 0$ the Hausdorff dimension of the set $\{x : \beta(x) = \delta\}$ if this set is non-empty and $-\infty$ otherwise (*cf.* [**19, 64**]). If the image of d is not discrete then it is said that f_α is a *multifractal function*. These concepts arised from a conjecture made by Frisch and Parisi [**33**] in the context of the study of turbulence. Examples of multifractal functions are scarce, and Jaffard showed in [**64**] that "Riemann's example" φ is indeed multifractal. Our last theorem establishes this result for any form that is not cuspidal, as the cuspidal case had already been resolved negatively in corollary 3.2 of [**19**].

THEOREM 3.6 (SPECTRUM OF SINGULARITIES). *Let d be the spectrum of singularities of either f_α, $\Re f_\alpha$ or $\Im f_\alpha$. Then:*

(i) If f is a cusp form:

$$d(\delta) = \begin{cases} 1 & \text{if } \delta = \alpha - r/2, \\ 0 & \text{if } \delta = 2\alpha - r, \\ -\infty & \text{otherwise.} \end{cases}$$

(ii) If f is not a cusp form:

$$d(\delta) = \begin{cases} 2 + 2\frac{\delta-\alpha}{r} & \text{if } \alpha - r \leq \delta \leq \alpha - r/2, \\ 0 & \text{if } \delta = 2\alpha - r \text{ and } f \text{ cuspidal at some rational}, \\ -\infty & \text{otherwise.} \end{cases}$$

The functions f_α, $\Re f_\alpha$ and $\Im f_\alpha$ are therefore multifractal if and only if f is not cuspidal.

The spectrum of singularities was known in the same cases as theorem 3.3. See [**19, 64**] and theorem 3.7 of [**83**].

3.3. Approximate functional equation

Our starting point is going to be an integral representation for f_α, given by the following lemma. This is exactly the Riemann-Liouville integral (**I**.10) we mentioned in the introduction.

LEMMA 3.7. *For $\alpha > \alpha_0$ the series* (3.2) *converges uniformly to a continuous function f_α, which admits the following integral representation*

$$f_\alpha(x) = \frac{(2\pi)^\alpha}{(im)^\alpha \Gamma(\alpha)} \int_{(x)} (z-x)^{\alpha-1} (f(z) - f(\infty))\, dz. \tag{3.4}$$

PROOF. Summing by parts (3.2) and using the estimates for partial sums given in proposition 2.10 it is clear that the series converges uniformly and hence to a continuous function. To prove the integral representation we start with

$$f_\alpha(x+iy) = \frac{(2\pi)^\alpha}{m^\alpha \Gamma(\alpha)} \int_0^\infty t^{\alpha-1} (f(x+iy+it) - f(\infty))\, dt, \tag{3.5}$$

identity that can be obtained from (3.1) integrating the series term by term because of the uniform convergence (with exponential decay) in the region $\Im z \geq y$. Now it suffices to take the limit $y \to 0^+$ on both sides. The left hand side corresponds to the Abel summation of a converging Fourier series, while in the right hand side the dominated convergence theorem applies with the bounds obtained in proposition 2.8. □

Having to deal with integrals of the kind (3.4) it is a natural question under which hypotheses we can apply the differentiation under the integral sign theorem. We prove here a particular version for convenience.

LEMMA 3.8. *Let $\gamma \in \mathrm{SL}_2(\mathbb{R})$ and let I be a bounded open interval whose closure does not contain the pole of γ. Let $g(z,x)$ be a function differentiable with respect to x in I and analytic for $z \in \mathbb{H}$. Assume moreover that both g and $g_x = \partial g/\partial x$ are jointly continuous, have exponential decay when $\Im z \to +\infty$ in vertical strips, uniformly in*

$x \in I$, and that for some $\beta > 0$, $\eta > 0$ they satisfy the following estimates when $z \to \gamma(x)$, also uniformly in $x \in I$:

$$g(z,x) = O\big((z-\gamma x)^{\beta+\eta-1}(\Im z)^{-\eta}\big),$$
$$g_x(z,x) = O\big((z-\gamma x)^{\beta+\eta-2}(\Im z)^{-\eta}\big).$$

Then the function

$$F(x) = \int_{(\gamma x)} g(z,x)\,dz \qquad \text{defined for} \quad x \in I$$

is in $\Lambda^\beta(I)$ for $0 < \beta < 1$, in $\Lambda^1_{\log}(I)$ for $\beta = 1$ and in $\mathcal{C}^{1,0}(I)$ for $\beta > 1$. In this last case,

$$F'(x) = \int_{(\gamma x)} g_x(z,x)\,dz \qquad \text{for} \quad x \in I.$$

Proof. Assume $x \in I$ and $0 < h < 1$ satisfying $x \pm h \in I$. Using Cauchy's theorem together with the estimates for g we can write for $0 < u < 1 < v$:

$$F(x \pm h) - F(x) = \int_{\gamma x + iu}^{\gamma x + iv} \big(g(z, x \pm h) - g(z,x)\big)\,dz + O\left(e^{-Kv} + u^\beta + hu^{\beta-1} + \frac{h^{\beta+\eta}}{u^\eta}\right).$$

It is clear now that F must be continuous, as for each ε we may choose u and v so that for h small enough $|F(x \pm h) - F(x)| \le \varepsilon$.

For the rest of the proof we choose $u = h$ and $v = +\infty$, so that the error term is of the form $O(h^\beta)$. By the mean value theorem:

$$|F(x \pm h) - F(x)| \ll h \int_{\gamma x + ih}^{\gamma x + i\infty} |g_x(z, x_z)|\;|dz| + O(h^\beta).$$

Using the estimates for g_x this last integral is of order $O(h^{\beta-1})$ for $0 < \beta < 1$ and of order $O(\log h)$ for $\beta = 1$.

Suppose now that $\beta > 1$. The estimates for g_x justify the use of the dominated convergence theorem, proving the existence and the formula for F'. Finally, the argument used to prove that F is continuous can be applied directly to F' substituting β by $\beta - 1$ to conclude that F' is also continuous. □

For the rest of this section we assume we are under the hypotheses of theorem 3.4, *i.e.*, σ is a fixed matrix in $\mathrm{SL}_2(\mathbb{R})$ whose bottom-left entry is negative and such that f^σ is a modular form for a finite index subgroup of $\mathrm{SL}_2(\mathbb{Z})$, x will denote an arbitrary real number different from $x_0 = \sigma\infty \in \mathbb{Q}$ and, for convenience, we also put $C_0 = (2\pi)^\alpha/((im)^\alpha\Gamma(\alpha))$.

To avoid unnecessary distractions we will hide some extra terms that appear during the subsequent manipulations inside the symbol $(\cdots)$; we will deal with them afterwards. The reader can check that all the missing terms appear in (3.6–3.9).

We start from lemma 3.7. Splitting the integral on the right hand side of (3.4) and performing the change of variables $z = \sigma w$ we have:

$$\begin{aligned} f_\alpha(x) &= C_0 \int_x^{x+2i} (z-x)^{\alpha-1} f(z)\,dz + (\cdots) \\ &= C_0 \int_S (\sigma w - x)^{\alpha-1} \big(j_\sigma(w)\big)^{r-2} \big(f^\sigma(w) - f(x_0)\big)\,dw + (\cdots). \end{aligned}$$

where S corresponds to a subarc of the geodesic halfcircle with endpoints $\sigma^{-1}(x)$ and $\sigma^{-1}(\infty)$, and $f(x_0) = \lim_{\Im z\to\infty} f^\sigma(z)$ (see the remarks after the statement of theorem 3.4). The integrand in the last equation has exponential decay when $\Im w \to +\infty$. This and the bounds from proposition 2.8 allow us to apply Cauchy's theorem to replace S with two vertical rays starting at the endpoints of S and projecting to $i\infty$:

$$f_\alpha(x) = C_0 \int_{(\sigma^{-1}x)} (\sigma w - x)^{\alpha-1} \big(j_\sigma(w)\big)^{r-2} \big(f^\sigma(w) - f(x_0)\big)\, dw + (\cdots).$$

By (*iii*) of proposition 2.1 we have the relation $(\sigma w - x) j_\sigma(w) = (w - \sigma^{-1}x) j_{\sigma^{-1}}(x)$. If we let C_1 denote the constant $C_0 e^{-2\pi i\alpha}$ if $x < x_0$ and C_0 otherwise, substituting:

$$f_\alpha(x) = C_1 \big(j_{\sigma^{-1}}(x)\big)^{\alpha-1} \int_{(\sigma^{-1}x)} (w-\sigma^{-1}x)^{\alpha-1} \big(j_\sigma(w)\big)^{r-\alpha-1} \big(f^\sigma(w)-f(x_0)\big)\, dw+(\cdots).$$

Let $\phi(w) = \big(j_\sigma(w)\big)^{r-\alpha-1}$ and denote by $\phi(\sigma^{-1}x^+)$ the limit of $\phi(w)$ when $w \to \sigma^{-1}x$ from the upper half-plane. Adding and subtracting $\phi(\sigma^{-1}x^+) = \big(j_{\sigma^{-1}}(x)\big)^{\alpha-r+1}$ and using that $j_{\sigma^{-1}}(x) = (-c)(x - x_0)$ where $c < 0$ is the lowest-left entry of σ, we arrive via lemma 3.7 to

$$f_\alpha(x) = B|x - x_0|^{2\alpha}(x - x_0)^{-r} f_\alpha^\sigma(\sigma^{-1}x) + (\cdots)$$

for $B = (m_{x_0}/m)^\alpha(-c)^{2\alpha-r} > 0$.

The terms we have omitted so far are the following ones:

(3.6)
$$(\cdots) = -C_0 \frac{(2i)^\alpha}{\alpha} f(\infty) + C_0 \int_{x+2i}^{x+i\infty} (z - x)^{\alpha-1} \big(f(z) - f(\infty)\big)\, dz$$

(3.7)
$$+ C_0 f(x_0) \int_x^{x+2i} (z - x)^{\alpha-1} \big(j_{\sigma^{-1}}(z)\big)^{-r}\, dz$$

(3.8)
$$+ C_0 \left(\int_{x_0}^{x_0+2i} + \int_{x_0+2i}^{x+2i} \right) (z - x)^{\alpha-1} \left(f(z) - \frac{f(x_0)}{\big(j_{\sigma^{-1}}(z)\big)^r} \right) dz$$

(3.9)
$$+ C(x - x_0)^{\alpha-1} \int_{(\sigma^{-1}x)} (w - \sigma^{-1}x)^{\alpha-1} \big(\phi(w) - \phi(\sigma^{-1}x^+)\big) \big(f^\sigma(w) - f(x_0)\big)\, dw.$$

The terms (3.6) and (3.8) make sense for any $x \in \mathbb{R}$ and are infinitely many times differentiable with respect to this variable. The other ones are studied in the following lemmas, which complete the proof of theorem 3.4:

Lemma 3.9. *The term* (3.7) *admits an expansion of the form:*

$$Ai^{-\alpha} f(x_0)\phi(x - x_0) + E(x)$$

where ϕ is defined in the statement of theorem 3.4. The constant A is real and nonzero and the error term $E(x)$ is infinitely many times differentiable.

Lemma 3.10. *The term* (3.9) *lies both in $\mathcal{C}^{1,0}(\mathbb{R} \setminus \{x_0\})$ and in the class $O(|x - x_0|^{2\alpha-r+1})$ when $x \to x_0$.*

PROOF OF LEMMA 3.9. We may assume that f is not cuspidal at x_0, since otherwise (3.7) is equal to zero. Note that in this case by hypothesis $\alpha > r$. Renaming $x - x_0$ to x if necessary we may further assume $x_0 = 0$. Hence up to a nonzero constant of the form $Ai^{-\alpha} f(x_0)$ we have to expand asymptotically the function

$$g(x) = \int_0^{2i} \frac{z^{\alpha-1}}{(x+z)^r}\, dz. \tag{3.10}$$

We will suppose for the moment that $0 < x < 1$ and $\alpha - r \notin \mathbb{Z}$. We have

$$g(x) = x^{-r} \int_0^{2xi} \frac{z^{\alpha-1}}{\left(1+\frac{z}{x}\right)^r}\, dz + \int_{2xi}^{2i} \frac{z^{\alpha-r-1}}{\left(1+\frac{x}{z}\right)^r}\, dz.$$

In the first integral we perform a linear change of variables, while in the second one we substitute the Laurent expansion

$$\left(1+\frac{x}{z}\right)^{-r} = \sum_{k\geq 0} \binom{-r}{k} x^k z^{-k}$$

which is uniformly convergent in the region $|z| \geq 2x$. Integrating term by term the expression now results

$$\begin{aligned} g(x) &= x^{\alpha-r} \int_0^{2i} \frac{z^{\alpha-1}}{(1+z)^r}\, dz + \sum_{k\geq 0} \binom{-r}{k} \frac{x^k}{\alpha-r-k} z^{\alpha-r-k} \Bigg|_{2xi}^{2i} \\ &= x^{\alpha-r} \left(\int_0^{2i} \frac{z^{\alpha-1}}{(1+z)^r}\, dz - \sum_{k\geq 0} \binom{-r}{k} \frac{(2i)^{\alpha-r-k}}{\alpha-r-k} \right) + h(x). \end{aligned} \tag{3.11}$$

where $h(x)$ is a function given by a power series which converges in a neighborhood of 0. Note that the expression within brackets is a constant A' satisfying

$$A' = \int_0^T \frac{z^{\alpha-1}}{(1+z)^r}\, dz - \sum_{k\geq 0} \binom{-r}{k} \frac{T^{\alpha-r-k}}{\alpha-r-k}$$

for any complex T with $|T| > 1$ and $\arg T \neq \pi$: the right hand side is indeed constant as can be easily checked by differentiating with respect to T. Hence

$$\begin{aligned} A' &= \lim_{T\to+\infty} \left(\int_0^T \frac{t^{\alpha-1}}{(1+t)^r}\, dt - \sum_{0\leq k<\alpha-r} \binom{-r}{k} \frac{T^{\alpha-r-k}}{\alpha-r-k} \right) \\ &= \int_0^\infty t^{\alpha-1} \left(\frac{1}{(1+t)^r} - \sum_{0\leq k<\alpha-r} \binom{-r}{k} \frac{1}{t^{r+k}} \right) dt. \end{aligned}$$

The sum corresponds to the Taylor expansion of order $\lfloor \alpha - r \rfloor$ of the function $(1+\xi)^{-r}$ multiplied by ξ^r and evaluated at $\xi = 1/t$. Since all the derivatives of this function have constant sign for $\xi > 0$ we deduce $A' \neq 0$. Although the exact value of A' is unimportant, using the integral formula for the error term in the Taylor expansion one can easily obtain a closed formula in terms of beta functions.

Suppose now that $\alpha - r$ is an integer. The same argument can be carried on, but when integrating the Laurent series term by term the term corresponding to $k = \alpha - r$ is now transformed into a logarithm. This term results

$$\binom{-r}{\alpha-r} x^{\alpha-r} \log z \Bigg|_{2xi}^{2i} = \binom{-r}{\alpha-r} x^{\alpha-r} \big(-\log(x/i) + \log 2 - \log T \big) \quad (T = 2i).$$

The first summand corresponds to the main term, while the other two should be merged into A'. This is relevant, as we will need $A' \in \mathbb{R}$ in order to handle the case $x < 0$. We may replace (3.11) with:

$$g(x) = -\binom{-r}{\alpha - r} x^{\alpha - r} \log(x/i) + A' x^{\alpha - r} + h(x). \tag{3.12}$$

Finally if $x < 0$, we go back to (3.10) and notice that

$$g(x) = (-1)^{\alpha - r} \overline{g(-x)},$$

and the very same equation is also satisfied by the main and error terms in equations (3.11) and (3.12). Therefore we may apply the results we have obtained for $x > 0$. □

Proof of lemma 3.10. Because of the extra cancelation as $w \to \sigma^{-1}x$ provided by the second factor inside the integral in (3.9) and the exponential decay given by the third factor when $\Im z \to +\infty$, lemma 3.8 can be applied with $\eta = \alpha_0$ and $\beta + \eta = \alpha + 1$. This shows that (3.9) is in $\mathcal{C}^{1,0}(\mathbb{R} \setminus \{x_0\})$ (and in fact it is possible to do a little better with a repeated application of the lemma).

For the second estimate, it suffices to show that

$$\int_{(\sigma^{-1}x)} (w - \sigma^{-1}x)^{\alpha - 1} (\phi(w) - \phi(\sigma^{-1}x^+)) (f^\sigma(w) - f(x_0))\, dw \ll |x - x_0|^{\alpha - r + 2} \tag{3.13}$$

when $x \to x_0$. Notice that for $w = \sigma^{-1}x + it$ we have

$$\phi(w) = \big(j_\sigma(w)\big)^{r - \alpha - 1} = \left(\frac{1}{(-c)(x - x_0)} + ict\right)^{r - \alpha - 1}$$

where c is the bottom-left entry of σ. Therefore applying the mean value theorem we obtain for $|x - x_0| \le 1$:

$$|\phi(w) - \phi(\sigma^{-1}x^+)| \ll \begin{cases} t|x - x_0|^{\alpha - r + 2} & t \le |x - x_0|^{-1} \\ t^{r - \alpha - 1} & t \ge |x - x_0|^{-1} \end{cases}.$$

We divide now the integration domain in three intervals and use these estimates, together with the trivial ones for f^σ, to show that the left hand side of (3.13) is

$$\begin{aligned} &\ll |x - x_0|^{\alpha - r + 2} \left(\int_0^1 t^\alpha (1 + t^{-\alpha_0})\, dt + \int_1^{|x - x_0|^{-1}} t^\alpha e^{-Kt}\, dt \right) \\ &\quad + \int_{|x - x_0|^{-1}}^\infty t^{r - 2} e^{-Kt}\, dt. \end{aligned}$$

This proves (3.13), since the first two integrals are convergent and the last one has exponential decay when $x \to x_0$. □

Proof of corollary 3.5. If f is a cusp form then (3.7) and the first summand of (3.6) vanish. Moreover since $\alpha = r - 1$ the function ϕ in (3.9) is constant, and hence this term also vanishes. The remaining terms are:

$$\begin{aligned} (\cdots) &= C_0 \left(\int_{x_0}^{x_0 + 2i} + \int_{x_0 + 2i}^{x + 2i} + \int_{x + 2i}^{x + i\infty} \right) (z - x)^{\alpha - 1} f(z)\, dz \\ &= \frac{(2\pi)^\alpha}{(im)^\alpha \Gamma(\alpha)} \int_{(x_0)} (z - x)^{\alpha - 1} f(z)\, dz. \end{aligned}$$

□

3.4. Wavelet transform

The wavelet transform was presented in the introduction (**I**.12) as the integral transform

$$Wf(a,b) = \frac{1}{a}\int_{\mathbb{R}} f(t)\,\bar{\psi}\left(\frac{t-b}{a}\right)\,dt \qquad \text{for} \quad a>0 \quad \text{and} \quad b\in\mathbb{R}.$$

where ψ has to be a wavelet, which we did not define rigorously. Actually, there is no unique definition: a wavelet can be any function which oscillates and at the same time has enough decay for the integral defining W to converge, and different definitions of the concept of wavelet can be found in the literature. In this dissertation we are going to stick to the following: given $\alpha>0$, a wavelet is a smooth function $\psi:\mathbb{R}\to\mathbb{C}$ satisfying:

(*i*) $\psi^{(k)}(x) \ll (1+|x|)^{-\alpha-1}$ for all $k\le\lceil\alpha\rceil$.

(*ii*) $\int_{\mathbb{R}} x^k\psi(x)\,dx = 0$ for $0\le k<\alpha$.

(*iii*) Either

$$\int_0^\infty |\hat{\psi}(\xi)|^2\,\frac{d\xi}{\xi} = \int_0^\infty |\hat{\psi}(-\xi)|^2\,\frac{d\xi}{\xi} = 1$$

or

$$\hat{\psi}(\xi)=0 \text{ if } \xi<0 \quad \text{and} \quad \int_0^\infty |\hat{\psi}(\xi)|^2\,\frac{d\xi}{\xi} = 1.$$

These axioms are adapted from the ones used by Jaffard [**65**, §2] to study "Riemann's example". The differences with the definition employed by Jaffard are subtle but important, and will allow us to avoid the very unnatural hypothesis that appear in the main theorems of the article [**19**].

Note (*i*) of the axioms implies that $\hat{\psi}$ exists and is ϵ-Hölder for some small ϵ, and together with (*ii*) this justifies the integrability of $|\hat{\psi}(\xi)|^2/\xi$. The decay of ψ also shows that Wf is well-defined for any bounded measurable function $f:\mathbb{R}\to\mathbb{C}$. If we moreover ask f to be continuous and periodic, with vanishing integral on each period, and satisfying $\hat{f}(\xi)=0$ for $\xi<0$ in the distributional sense[4] in case the same is satisfied by ψ, then the following inversion formula holds:

$$f(x) = \int_{\mathbb{R}^+}\int_{\mathbb{R}} Wf(a,b)\psi\left(\frac{x-b}{a}\right)\,\frac{db\,da}{a^2}. \tag{3.14}$$

The proof of the inversion formula can be found in [**52**] with weaker hypotheses, but nevertheless we will provide an adapted version here for convenience. The outer integral in (3.14) in principle has to be understood as an improper Riemann integral, but in our applications it will be absolutely convergent.

The wavelet transform allows us to reformulate questions concerning the regularity of f in a point x_0 as questions about the growth of its wavelet transform W in a neighborhood of the corresponding point $(0^+,x_0)$, as it is shown by the following two theorems:

THEOREM 3.11. *Let $0<\beta<\alpha$ and f as above. If $f\in\mathcal{C}^\beta(x_0)$ then*

$$Wf(a,b) \ll a^\beta + |b-x_0|^\beta$$

when $(a,b)\to(0^+,x_0)$.

[4]This means that whenever ϕ is a compactly supported (or of fast decay) smooth function whose support is contained in $\{x<0\}$ we have $\int f(\xi)\hat{\phi}(\xi)\,d\xi = 0$. See §3.8 of [**12**].

THEOREM 3.12. *Let $0 < \beta' < \beta < \alpha$ and f as above. If*

$$Wf(a,b) \ll a^\beta + a^{\beta-\beta'}|b-x_0|^{\beta'}$$

when $(a,b) \to (0^+, x_0)$ then $f \in \mathcal{C}^\beta(x_0)$ if β is not an integer and $f \in \mathcal{C}^\beta_{\log}(x_0)$ otherwise.

The bounds involving $Wf(a,b)$ in these two theorems may also be written in the forms $a^\beta\left(1+\frac{|b-x_0|}{a}\right)^\beta$ and $a^\beta\left(1+\frac{|b-x_0|}{a}\right)^{\beta'}$, respectively, from where it is clear that the second one constitutes a strengthening of the first.

REMARK. *The last two theorems are adapted from proposition 1 of Jaffard's article* [**65**] *for our definition of wavelet. With our notation, the use of the definition employed by Jaffard would require the extra hypothesis $\lfloor\beta\rfloor \le \lfloor\alpha\rfloor - 1$ in the theorems, which was the problem encountered in* [**19**]*. Note also that the logarithm appearing in theorem 3.12 when β is an integer is neglected in* [**65**] *(and in fact, the proof for $\beta \ge 1$ left to the reader). Indeed, the approximate functional equation (theorem 3.4) shows that the logarithm may very well be necessary for some functions satisfying the hypotheses (*cf. *§3.7).*

PROOF OF THE INVERSION FORMULA. (Adapted from [**52**]) Assume first we are in the first case of axiom (*iii*). Let $\epsilon > 0$ and

$$g_\epsilon(x) = \int_\epsilon^{1/\epsilon}\int_{\mathbb{R}} Wf(a,b)\psi\left(\frac{x-b}{a}\right)\frac{db\,da}{a^2}.$$

We must show $\lim_{\epsilon\to 0^+} g_\epsilon(x) = f(x)$. Substituting the definition of Wf and applying Fubini twice,

$$g_\epsilon(x) = \int_{-\infty}^{+\infty} f(t)\int_\epsilon^{1/\epsilon}\frac{1}{a^3}\int_{-\infty}^{+\infty}\bar\psi\left(\frac{t-b}{a}\right)\psi\left(\frac{x-b}{a}\right)db\,da\,dt. \tag{3.15}$$

The change of variables $(x-b)/a \mapsto b$ in the inner integral shows

$$g_\epsilon(x) = \int_{-\infty}^{+\infty} f(t)\int_\epsilon^{1/\epsilon}\frac{1}{a^2}h\left(\frac{t-x}{a}\right)da\,dt$$

where $h(t) = \int_{-\infty}^{+\infty}\bar\psi(t+b)\psi(b)\,db$. We perform now the change of variables $(t-x)/a \mapsto a$, obtaining

$$\begin{aligned} g_\epsilon(x) &= \int_{-\infty}^{+\infty}\frac{f(t)}{t-x}\int_{\epsilon(t-x)}^{(t-x)/\epsilon} h(a)\,da\,dt \\ &= \int_{-\infty}^{+\infty} f(t)\left(\frac{1}{\epsilon}M\big((t-x)/\epsilon\big) - \epsilon M\big(\epsilon(t-x)\big)\right)dt \end{aligned}$$

for $M(t) = t^{-1}\int_0^t h(\tau)\,d\tau$. We claim $M \in L^1(\mathbb{R})$ and $\int M = 1$. If so, using that f is continuous and periodic with vanishing integral in each period,

$$\lim_{\epsilon\to 0^+}\frac{1}{\epsilon}\int_{-\infty}^{+\infty} f(t)M\big((t-x)/\epsilon\big)\,dt = f(x) \quad \text{and} \quad \lim_{\epsilon\to 0^+}\epsilon\int_{-\infty}^{+\infty} f(t)M\big(\epsilon(t-x)\big)\,dt = 0,$$

the first equality because $\epsilon^{-1}M(t/\epsilon)$ is an approximation of the identity and the second because of a Riemann-Lebesgue lemma adapted for f, or directly integrating by parts since M is smooth. Hence $\lim_{\epsilon\to 0^+} g_\epsilon(x) = f(x)$.

To prove the claim we first need show

$$\int_{-\infty}^0 h(\tau)\,d\tau = \int_0^\infty h(\tau)\,d\tau = 0.$$

Note from the definition of h that $h(t) \ll (1+|t|)^{-\alpha-1}$, and hence $h \in L^1(\mathbb{R})$. Also, by the Plancherel formula, $h(t) = \int_{\mathbb{R}} e(-t\xi)|\hat{\psi}(\xi)|^2\,d\xi$. Therefore

$$\int_0^t h(\tau)\,d\tau = \int_0^t \int_{-\infty}^{+\infty} e(-\tau\xi)|\hat{\psi}(\xi)|^2\,d\xi\,d\tau = -\frac{1}{2\pi i}\int_{-\infty}^{+\infty} |\hat{\psi}(\xi)|^2 e(-t\xi)\,\frac{d\xi}{\xi},$$

where we have used Fubini and (*iii*) of the wavelet axioms. By the Riemann-Lebesgue lemma this vanishes when $t \to \pm\infty$.

Hence for $t > 0$ (resp. $t < 0$) we have $M(t) = -t^{-1}\int_t^{+\infty} h(\tau)\,d\tau$ (resp. $M(t) = t^{-1}\int_{-\infty}^t h(\tau)\,d\tau$) and therefore $M(t) \ll (1+|t|)^{-\alpha-1}$, implying $M \in L^1(\mathbb{R})$. For any $\epsilon > 0$ consider $M_\epsilon(t) = (t - i\epsilon)^{-1}\int_0^t h(\tau)\,d\tau$, and apply Fubini to write

$$\int_{-T}^{T} M_\epsilon(t)\,dt = -\frac{1}{2\pi i}\int_{-\infty}^{+\infty} |\hat{\psi}(\xi)|^2 \int_{-T}^{T} \frac{e(-t\xi)}{t - i\epsilon}\,dt\,\frac{d\xi}{\xi}.$$

An application of Cauchy's theorem and a direct estimation shows that the inner integral equals $2\pi i H(-\xi)e^{2\pi\epsilon\xi} + O\big(\min(|\xi T|^{-1}, \log(T/\epsilon))\big)$ where H is the Heaviside function: $H(\xi) = 1$ for $\xi > 0$ and $H(\xi) = 0$ for $\xi < 0$. Substituting and carefully taking the limit first when $T \to \infty$ and then when $\epsilon \to 0$ we obtain

$$\int_{-\infty}^{+\infty} M(t)\,dt = \int_0^{\infty} |\hat{\psi}(-\xi)|^2\,\frac{d\xi}{\xi} = 1.$$

Finally suppose that the wavelet ψ satisfies instead the second part of axiom (*iii*). Then $\psi + \bar{\psi}$ is a wavelet satisfying the first part of (*iii*) and therefore the inversion formula holds for it. But Wf has the same values with respect to both wavelets, and the same is true for the inner integral in (3.15). This essentially follows from the Plancherel formula, but in the first case to avoid working with the distribution $\hat{f}$ it is convenient to apply directly the definition of $\operatorname{supp}\hat{f} \subset \{x \geq 0\}$ with some smoothing and truncation (see footnote 4). In fact, $\int f\hat{\phi} = 0$ for any continuous function $\phi \in L^1(\mathbb{R})$ supported in $\{x \geq 0\}$ and satisfying $\hat{\phi} \in L^1(\mathbb{R})$. □

Proof of theorem 3.11. We can assume without loss of generality $x_0 = 0$. By hypothesis there is a polynomial P of degree strictly smaller than β such that

$$|f(x) - P(x)| \ll |x|^\beta,$$

estimate which we may assume to hold globally. Hence, by the axioms (*i*) and (*ii*) of the wavelet definition,

$$\begin{aligned} Wf(a,b) &\ll \frac{1}{a}\int_{\mathbb{R}} |f(t) - P(t)|\left|\psi\left(\frac{t-b}{a}\right)\right| dt \\ &\ll \frac{1}{a}\int_{\mathbb{R}} \frac{|t|^\beta}{\left(\left|\frac{t-b}{a}\right| + 1\right)^{\alpha+1}}\,dt \\ &\ll a^\beta \int_{\mathbb{R}} \frac{|t|^\beta}{(|t|+1)^{\alpha+1}}\,dt + |b|^\beta \int_{\mathbb{R}} \frac{dt}{(|t|+1)^{\alpha+1}} \\ &\ll a^\beta + |b|^\beta. \end{aligned}$$

□

In order to prove theorem 3.12 we shall use the inversion formula (3.14), which for convenience will be written in the following way:

$$f(x) = \int_{\mathbb{R}^+} \omega(a,x)\,\frac{da}{a} \tag{3.16}$$

where

$$\omega(a,x) = \frac{1}{a}\int_{\mathbb{R}} Wf(a,b)\psi\left(\frac{x-b}{a}\right)\,db. \tag{3.17}$$

We prove first some estimates for ω. In particular they show that the integral in (3.16) is absolutely convergent for x sufficiently close to x_0.

LEMMA 3.13. *Under the hypotheses of theorem 3.12 the function $x \mapsto \omega(a,x)$ is infinitely many times differentiable and satisfies for all $k \leq \lceil \alpha \rceil$ and for some $\delta > 0$:*

$$\frac{\partial^k \omega}{\partial x^k}(a,x) \ll a^{-k-1}, \tag{3.18}$$

$$\frac{\partial^k \omega}{\partial x^k}(a,x) \ll a^{\beta-k} + a^{\beta-\beta'-k}|x-x_0|^{\beta'} \qquad \text{for} \quad a \leq 1,\ |x-x_0| \leq \delta. \tag{3.19}$$

PROOF. It is clear that $Wf(a,b)$ is uniformly bounded and ψ and all its derivatives up to $\lceil \alpha \rceil$ have decay (axiom (i)). Therefore we may differentiate (3.17) under the integral sign obtaining

$$\frac{\partial^k \omega}{\partial x^k}(a,x) = \frac{1}{a^{k+1}}\int_{\mathbb{R}} Wf(a,b)\psi^{(k)}\left(\frac{x-b}{a}\right)\,db. \tag{3.20}$$

Integrating by parts in the definition of $Wf(a,b)$ and using that the integral over each period of f vanishes it is readily seen that $Wf(a,b) \ll a^{-1}$. Plugging this into (3.20) one obtains (3.18).

To prove (3.19) we first assume without loss of generality that $x_0 = 0$, and that the bounds in the statement of theorem 3.12 hold uniformly in the neighborhood $a \leq 1$ and $|b| \leq 2\delta$. We have for $a \leq 1$ and $|x| \leq \delta$:

$$\begin{aligned}
\frac{\partial^k \omega}{\partial x^k}(a,x) &\ll \frac{1}{a^{k+1}}\int_{|b|\leq 2\delta} \frac{a^\beta + a^{\beta-\beta'}|b|^{\beta'}}{\left(\left|\frac{x-b}{a}\right|+1\right)^{\alpha+1}}\,db + \frac{1}{a^{k+1}}\int_{|b|>2\delta} \frac{db}{\left(\left|\frac{x-b}{a}\right|+1\right)^{\alpha+1}} \\
&\ll a^{\beta-k} + a^{\beta-\beta'-k}\int_{\mathbb{R}} \frac{|x-at|^{\beta'}}{(|t|+1)^{\alpha+1}}\,dt + \frac{1}{a^k}\int_{t>\delta/a} \frac{dt}{(t+1)^{\alpha+1}} \\
&\ll a^{\beta-k} + a^{\beta-\beta'-k}|x|^{\beta'}.
\end{aligned}$$

□

PROOF OF THEOREM 3.12. Again we can assume $x_0 = 0$. Let $N = \lfloor \beta \rfloor$ if β is not an integer and $N = \beta - 1$ otherwise, *i.e.* $N = \lceil \beta \rceil - 1$. We perform a Taylor expansion of order N on ω:

$$\omega(a,x) = \sum_{k=0}^{N} \frac{\partial^k \omega}{\partial x^k}(a,0)\frac{x^k}{k!} + E(a,x).$$

Using the bounds of lemma 3.13 we can plug this into (3.16) to obtain

$$f(x) = P(x) + \int_{\mathbb{R}^+} E(a,x)\,\frac{da}{a}$$

for certain polynomial P of degree $N < \beta$. It suffices to prove that the integral term has the right behavior when $x \to 0$.

We split the integral. In the range $a \leq |x|$ we use (3.19) with either $x = 0$ or $k = 0$ to obtain

$$\left|\int_{a\leq|x|} E(a,x)\,\frac{da}{a}\right| \leq \int_{a\leq|x|} |\omega(a,x)|\,\frac{da}{a} + \sum_{k=0}^{N}\frac{|x|^k}{k!}\int_{a\leq|x|}\left|\frac{\partial^k\omega}{\partial x^k}(a,0)\right|\frac{da}{a} \ll |x|^\beta.$$

In the complementary range, assuming that β is not an integer, we use the formula for the Taylor error term together with (3.19):

$$\left|\int_{a\geq|x|} E(a,x)\,\frac{da}{a}\right| \leq \frac{|x|^{N+1}}{(N+1)!}\int_{a\geq|x|}\left|\frac{\partial^{N+1}\omega}{\partial x^{N+1}}(a,\xi_{a,x})\right|\frac{da}{a} \ll |x|^{\beta}.$$

When β is an integer the same argument works using (3.19) in the range $|x| \leq a \leq 1$ and (3.18) in the range $a \geq 1$. The right hand side has to be replaced by $|x|^{\beta}\log|x|$. □

Following [**19, 65**] we apply these theorems to f_α, where f is a modular form, with $\psi(x) = (x+i)^{-\alpha-1}$. The reader can easily verify that ψ satisfies the axioms (i) and (ii) of our definition of wavelet. In order to check axiom (iii) we compute $\hat{\psi}$. The integral

$$\hat{\psi}(\xi) = \int_{\mathbb{R}} \frac{e^{-2\pi i\xi x}}{(x+i)^{\alpha+1}}\,dx$$

vanishes for $\xi \leq 0$ by Cauchy's theorem. For $\xi > 0$ we perform a change of variables obtaining

$$\hat{\psi}(\xi) = \xi^{\alpha}e^{-2\pi\xi}\int_{\mathbb{R}+\xi i}\frac{e^{-2\pi iz}}{z^{\alpha+1}}\,dz$$

and by Cauchy's theorem the integral on the right hand side is a constant with respect to ξ. The exact value of the constant is not important, since ψ need not be normalized for theorems 3.11 and 3.12 to hold, although it can be explicitly computed by means of Hankel's contour integral for the reciprocal of the gamma function (*cf.* [**96**, §12.22]).

It is also clear that f_α is a periodic function (since we have assumed $\kappa_\infty \in \mathbb{Q}$ *cf.* §2.5), with vanishing integral on each period, and whose Fourier transform (in the distributional sense) is supported only in the positive frequencies. To compute its wavelet transform with respect to ψ it suffices to compute the one for $g(x) = e^{2\pi i\lambda x}$. This can be done using some basic properties of the Fourier transform:

$$Wg(a,b) = e^{2\pi i\lambda b}\bar{\hat{\psi}}(\lambda a) = \begin{cases} Ca^{\alpha}\lambda^{\alpha}e^{2\pi i\lambda(b+ai)} & \lambda > 0 \\ 0 & \lambda \leq 0. \end{cases} \tag{3.21}$$

Hence

$$Wf_\alpha(a,b) = C'a^{\alpha}(f(b+ai) - f(\infty)). \tag{3.22}$$

Corollary 3.14. *If for some $0 < \beta < \alpha$ one has $f_\alpha \in \mathcal{C}^{\beta}(x_0)$ then*

$$f(b+ai) \ll a^{\beta-\alpha} + a^{-\alpha}|b-x_0|^{\beta}$$

when $(a,b) \to (0^+, x_0)$. Reciprocally, if for some $0 < \beta' < \beta < \alpha$ one has

$$f(b+ai) \ll a^{\beta-\alpha} + a^{\beta-\beta'-\alpha}|b-x_0|^{\beta'}$$

when $(a,b) \to (0^+, x_0)$, then $f_\alpha \in \mathcal{C}^{\beta}(x_0)$ if β is not an integer and $f_\alpha \in \mathcal{C}^{\beta}_{\log}(x_0)$ otherwise. Moreover both statements remain true if one replaces f_α by its real or imaginary parts.

Proof. The part of the theorem concerning f_α follows at once from theorems 3.11 and 3.12 and (3.22). Also note that if $f_\alpha \in \mathcal{C}^{\beta}(x_0)$ or $f_\alpha \in \mathcal{C}^{\beta}_{\log}(x_0)$ then the same must hold for the real and the imaginary parts of f_α.

On the other hand, $\Re f_\alpha$ and $\Im f_\alpha$ are bounded functions, and hence their wavelet transforms are well defined. By rewriting the sine and cosine functions involved in their Fourier series as sums of exponentials and applying (3.21) one obtains

$$Wf_\alpha(a,b) = 2W\Re f_\alpha(a,b) = 2iW\Im f_\alpha(a,b).$$

Since the inversion formula (3.14) was not used in the proof of theorem 3.11, we may apply this theorem to $\Re f_\alpha$ and $\Im f_\alpha$. □

3.5. Proof of the regularity theorems

This section contains the proofs of theorems 3.1, 3.2 and 3.3.

LEMMA 3.15. *For any integer $k < \alpha - \alpha_0$ we have $f_\alpha \in \mathcal{C}^{k,0}(\mathbb{R})$ and $f_\alpha^{(k)} = (2\pi i/m)^k f_{\alpha-k}$. If moreover f_α cannot be continuously differentiated $k+1$ times in any open interval containing a point x, then*

$$\beta^*(x) = k + \min\left(1, \beta_{\alpha-k}(x)\right),$$

where $\beta_{\alpha-k}$ denotes the pointwise Hölder exponent of $f_{\alpha-k}$. This formula extends to $\Re f_\alpha$ and $\Im f_\alpha$ if both these functions satisfy the nondifferentiability hypothesis and their pointwise Hölder exponents coincide.

PROOF. Since $\alpha - k > \alpha_0$ the series defining $f_{\alpha-k}$ converges uniformly (lemma 3.7), and therefore can be integrated term by term. This shows $f_\alpha \in \mathcal{C}^{k,0}(\mathbb{R})$ and that the formula for $f_\alpha^{(k)}$ holds. The rest follows from the definition of β^*. □

In order to prove theorems 3.1 and 3.2 we anticipate two very simple facts which will come in handy. Applying corollary 3.14 with the bounds from proposition 2.8 we obtain $\beta(x) = \alpha - r/2$ for f cuspidal and x irrational and $\beta(x) = \alpha - r$ for f not cuspidal and x any non-cuspidal rational. In the rest of cases, $\beta(x) \geq \alpha - \alpha_0$. The same results hold for the pointwise Hölder exponent of both $\Re f_\alpha$ and $\Im f_\alpha$.

PROOF OF THEOREM 3.1. (*i*) (Adapted from proposition 3.1 of [**14**]) If the series defining f_α converges at a certain point for $\alpha < \alpha_0$ then summing by parts the series defining f_{α_0} must also converge at that point, and therefore we may reduce to this case.

Suppose first that f is cuspidal, we will prove that $f_{r/2}$ diverges at any irrational point x. We can assume, rescaling f, that $m = 1$ and $\kappa = 0$. Considering the kernels of summability $\varphi_1(u) = e^{-2\pi u}(u^{r/2}+1)$ and $\varphi_2(u) = e^{-2\pi u}$ (see §**A**.3), we have:

$$\lim_{y\to 0^+} y^{r/2} f(x+iy) = \lim_{y\to 0^+} \left(\sum_{n>0} A_n \varphi_1(ny) - \sum_{n>0} A_n \varphi_2(ny) \right) = 0$$

with $A_n = \frac{a_n}{n^{r/2}} e^{2\pi i n x}$, as long as $f_{r/2}$ converges at x; but this contradicts proposition 2.8.

Suppose now that f is not cuspidal. We prove that f_r is not Abel summable at any non-cuspidal rational point x. If this were not the case then by (3.5) of lemma 3.7 we would have for some $\ell \in \mathbb{C}$,

$$\ell = \lim_{y\to 0^+} f_r(x+iy) = \lim_{y\to 0^+} \frac{(2\pi)^r}{m^r \Gamma(r)} \int_y^\infty (t-y)^{r-1} \big(f(x+it) - f(\infty)\big)\, dt.$$

But since by the expansion at the cusp the term $f(x+it)$ behaves like Ct^{-r} for small t, the right hand side diverges.

(*ii*) By lemma 3.15 the function f_α is continuously differentiable k times, where $k = \lfloor \alpha - \alpha_0 \rfloor$ if $\alpha - \alpha_0$ is not an integer and $k = \alpha - \alpha_0 - 1$ otherwise. The result now follows from applying lemma 3.8 to the integral representation given by lemma 3.7 for $f_{\alpha-k}$.

(*iii*) Suppose first that f is not cuspidal. If $\alpha - r < 1$ then neither f_α nor its real or nor its imaginary part are differentiable at any non-cuspidal rational, since they are at most $(\alpha - r)$-Hölder at these points. Only the limit case $\alpha = r + 1$ remains. But in this case we may appeal to theorem 3.4, since $2\alpha - r = r + 2 > 1$ implies that both the second term and the error term are differentiable at the rational x_0, and the first term is not if x_0 is non-cuspidal. A more detailed analysis shows that neither the real nor the imaginary parts of the function $Cx \log x$ are differentiable at 0 for any complex constant C, and hence the same applies for both $\Re f_\alpha$ and $\Im f_\alpha$.

(Adapted from lemma 3.7 of [**19**]) Suppose now that f is cuspidal, and rescaling $m = 1$ and $\kappa = 0$. If f_α is in $\mathcal{C}^{1,0}(I)$ then by theorem 3.4 it is also in $\mathcal{C}^{1,0}(\gamma(I))$ for any $\gamma \in \Gamma$. It follows that f'_α must exist and be continuous everywhere, for example by choosing γ with the pole inside I so that $\gamma(I)$ covers a whole period of f_α. This is possible because the equivalence class $[\infty]$ is dense (proposition 2.3). Integrating by parts in $\int_0^1 f'_\alpha(x) e(-nx)\, dx$ to obtain the Fourier coefficients of f'_α and using Bessel's inequality,

$$\|f'_\alpha\|^2 \gg \sum_{n>0} \frac{|a_n|^2}{n^{2\alpha-2}}.$$

But the right hand side diverges for $\alpha - r/2 \leq 1$ as can be checked by summing by parts and using the estimates of proposition 2.11.

Finally assume that either $\Re f_\alpha$ or $\Im f_\alpha$ is in $\mathcal{C}^{1,0}(I)$. Since the constant B in theorem 3.4 is real, the same argument works as long as we can find $\gamma \in \Gamma$ with the pole in I and $\mu_\gamma \in \{\pm 1\}$. One such matrix can be constructed as follows: pick a rational number $x \in I$ and let $\eta \in \Gamma$ be a parabolic matrix fixing x of positive trace with negative bottom-left entry. Since $\lim_{n\to\infty} \eta^{-n}\infty = x$, for n big enough η^n has its pole inside I. Moreover μ_η is a root of unity and $\mu_{\eta^n} = \mu_\eta^n$ (see §2.5). Hence we can choose $\gamma = \eta^n$ for a carefully chosen n. □

Proof of theorem 3.2. Let x_0 be a rational number.

(*i*) If f is not cuspidal at x_0 then we already know $\beta(x_0) = \alpha - r$. Hence may assume that f is cuspidal at x_0. Choose a matrix $\sigma \in \mathrm{SL}_2(\mathbb{Z})$ with negative bottom-left entry satisfying $\sigma\infty = x_0$ and apply theorem 3.4. We deduce that $f_\alpha \in \mathcal{C}^{2\alpha-r}(x_0)$ and that $f_\alpha \notin \mathcal{C}^{2\alpha-r+\varepsilon}(x_0)$ for any $\varepsilon > 0$, since the term $\sigma^{-1}x$ diverges to ∞ when $x \to x_0$ and f_α^σ is a nonconstant periodic function. Hence $\beta(x_0) = 2\alpha - r$. The same must be true for $\Re f_\alpha$ and $\Im f_\alpha$ as long as neither $\Re f_\alpha^\sigma$ nor $\Im f_\alpha^\sigma$ are constants. This is indeed the case as f_α^σ corresponds to a Fourier series with only positive frequencies.

(*ii*) The exponent β^* is determined by applying lemma 3.15 with $k = [\alpha - \alpha_0]$ if $\alpha - \alpha_0 \notin \mathbb{Z}$ and $k = \alpha - \alpha_0 - 1$ otherwise (*cf.* theorem 3.1).

(*iii*) To determine β^{**} note first that theorem 3.1 implies $\beta^{**}(x) \geq \alpha - \alpha_0$. Since this exponent also satisfies $\beta^{**}(x) \leq \liminf_{t\to x} \beta(t)$, as can be readily seen from its definition, and we have $\beta(x) = \alpha - \alpha_0$ for a dense set (the irrational numbers if f is cuspidal and the non-cuspidal rationals otherwise) we conclude $\beta^{**}(x) = \alpha - \alpha_0$ for all x.

(*iv*) The case x_0 non-cuspidal has already been treated in the proof of theorem 3.1, part (*iii*). Hence we may suppose that f is cuspidal at x_0. We appeal again

to theorem 3.4 but now we will use the explicit expression for the error term (*cf.* §3.3):

$$f_\alpha(x) = B|x-x_0|^{2\alpha}(x-x_0)^{-r} f_\alpha^\sigma(\sigma^{-1}x) + (3.6) + (3.8) + (3.9).$$

Terms (3.6) and (3.8) are everywhere differentiable, while term (3.9) can be differentiated at x_0 by lemma 3.10. Hence f_α is differentiable at x_0 if and only if the first summand is. Since f_α^σ is bounded, nonconstant and periodic this will happen if and only if $2\alpha - r > 1$. The same must be true for the real and imaginary parts of f_α, as neither $\Re f_\alpha^\sigma$ nor $\Im f_\alpha^\sigma$ are constants.

Hence whenever $f'_\alpha(x_0)$ exists it is given by the sum of the derivatives of the terms (3.6) and (3.8) evaluated at x_0 (the other terms have vanishing derivative at x_0). Differentiating under the integral sign and integrating by parts one obtains the desired formula. □

Proof of theorem 3.3. Let x_0 an irrational number. The pointwise Hölder exponent $\beta(x_0)$ is deduced by applying corollary 3.14 to the estimates of proposition 2.8 if f is a cusp form (see remark above) and of proposition 2.12 otherwise. The exponent $\beta^*(x_0)$ follows from lemma 3.15, while $\beta^{**}(x_0)$ was already determined in the proof of theorem 3.2, part (*iii*). □

3.6. Spectrum of singularities

In order to prove theorem 3.6 we will need some tools from Diophantine approximation theory. More concretely we will need a refinement of the following classic theorem:

Theorem 3.16 (Jarník-Besicovitch). *Let $\tau \geq 2$. The Hausdorff dimension of the set*

$$A_\tau := \left\{ x \,:\, \left| x - \frac{p}{q} \right| \ll \frac{1}{q^\tau} \text{ for infinitely many rationals } \frac{p}{q} \right\}$$

is $2/\tau$. Moreover, if we denote by $\mathcal{H}^t$ the t-dimensional outer Hausdorff measure, $\mathcal{H}^{2/\tau}(A_\tau) = \infty$.

For the proof of theorem 3.16 when $\tau > 2$ we refer the reader to Jarnik's original paper [**66**].[5] The theorem appearing there corresponds to the stronger Diophantine condition $|x - p/q| \leq q^{-\tau}$, but the result can be readily translated to our statement. The case $\tau = 2$ follows from Dirichlet's approximation theorem.

Throughout this section we are going to reserve the bold letters $\mathfrak{a}, \mathfrak{b}, \ldots$ to denote cusps, and we are going to write $\mathfrak{a} \sim \mathfrak{b}$ to denote that these two cusps lie in the same orbit modulo Γ, *i.e.*, that $[\mathfrak{a}] = [\mathfrak{b}]$ or $\mathfrak{b} = \gamma(\mathfrak{a})$ for some $\gamma \in \Gamma$. The theorem we need is the following, which takes into account that rational numbers are well distributed among the different classes of cusps.

Theorem 3.17. *Let $\mathfrak{a}$ be a cusp for Γ and $\tau \geq 2$. The Hausdorff dimension of the set*

$$A_\tau^{\mathfrak{a}} := \left\{ x \,:\, \left| x - \frac{p}{q} \right| \ll \frac{1}{q^\tau} \text{ for infinitely many rationals } \frac{p}{q} \sim \mathfrak{a} \right\}$$

is $2/\tau$. Moreover, if we denote by $\mathcal{H}^t$ the t-dimensional outer Hausdorff measure, $\mathcal{H}^{2/\tau}(A_\tau^{\mathfrak{a}}) = \infty$.

Theorem 3.17 is a particular case of more general results about Fuchsian groups (*cf.* [**91**]). We provide here an elementary proof based on theorem 3.16.

[5] See [**7**] for a survey in English.

PROOF. Note that we may assume without loss of generality that Γ is a normal subgroup of $\mathrm{SL}_2(\mathbb{Z})$. Indeed, if this is not the case, we simply replace Γ with the biggest normal group it contains, *i.e.*, the intersection of all its conjugates. The normality of Γ implies that the action of $\mathrm{SL}_2(\mathbb{Z})$ on the equivalence classes of cusps modulo Γ is well-defined.

Let γ be any matrix in $\mathrm{SL}_2(\mathbb{Z})$ and x an irrational number in $A_\tau^{\mathfrak{a}}$. We claim that if p/q is a rational number in a neighborhood of x and q' denotes the denominator of $\gamma(p/q)$ then $q' \ll q$, the implicit constant depending on x and γ. Indeed $q' = cp+dq$, and $p \ll q$ because $|p/q| \sim |x|$. From this together with the mean value theorem applied to $|\gamma(x) - \gamma(p/q)|$ we deduce that $\gamma(x) \in A_\tau^{\gamma(\mathfrak{a})}$. The argument can also be applied to γ^{-1} and therefore:

$$\gamma(A_\tau^{\mathfrak{a}}) = A_\tau^{\gamma(\mathfrak{a})}. \tag{3.23}$$

For any Lipschitz function h with Lipschitz constant C and any set Ω we have

$$\mathcal{H}^t\big(h(\Omega)\big) \leq C^t \mathcal{H}^t(\Omega). \tag{3.24}$$

This follows from the definition of Hausdorff outer measure. We want to apply this to prove that all the sets $A_\tau^{\mathfrak{a}}$ have roughly the same size when $\mathfrak{a}$ ranges through a set of representatives of the equivalence classes of the cusps modulo Γ, but the Möbius transformation γ is not Lipschitz in any neighborhood of its pole. This problem has a simple workaround. Let m be the width of the cusp ∞ and I any interval of length m not containing the pole of γ, and whose image $J = \gamma(I)$ is also of length m. Then from (3.23) we have

$$\begin{aligned} \gamma(A_\tau^{\mathfrak{a}} \cap I) &= A_\tau^{\gamma(\mathfrak{a})} \cap J \\ A_\tau^{\mathfrak{a}} + m &= A_\tau^{\mathfrak{a}}. \end{aligned}$$

Applying (3.24),

$$\mathcal{H}^t(A_\tau^{\gamma(\mathfrak{a})}) \ll \mathcal{H}^t(A_\tau^{\mathfrak{a}}).$$

The opposite inequality is also true and hence the Hausdorff dimension of the set $A_\tau^{\mathfrak{a}}$ must be independent of $\mathfrak{a}$. Since we also know by theorem 3.16 that $A_\tau = \bigcup_{\mathfrak{a}} A_\tau^{\mathfrak{a}}$ has dimension $2/\tau$, we conclude that all the $A_\tau^{\mathfrak{a}}$ must have exactly that dimension. It is also immediate that $\mathcal{H}^{2/\tau}(A_\tau^{\mathfrak{a}}) = \infty$. □

COROLLARY 3.18. *Let $2 \leq \tau \leq \infty$. The Hausdorff dimension of the set $\{x : \tau_x = \tau\}$ is $2/\tau$.*

For the definition of τ_x see (3.3).

PROOF. Assume $\tau > 2$ and let Ξ be a set of representatives of the equivalence classes of cusps at which f is not cuspidal. We have the identity

$$\{x : \tau_x = \tau\} = \bigcap_{\tau' < \tau} \bigcup_{\mathfrak{a} \in \Xi} A_{\tau'}^{\mathfrak{a}} \setminus \bigcup_{\tau' > \tau} \bigcup_{\mathfrak{a} \in \Xi} A_{\tau'}^{\mathfrak{a}}.$$

By theorem 3.17 the set on the right hand side has Hausdorff dimension at most $2/\tau$. On the other hand from the same theorem one deduces that for $\tau < +\infty$ we have

$$\mathcal{H}^{2/\tau}\Big(\bigcap_{\tau' < \tau} \bigcup_{\mathfrak{a} \in \Xi} A_{\tau'}^{\mathfrak{a}} \Big) = \infty, \qquad \mathcal{H}^{2/\tau}\Big(\bigcup_{\tau' > \tau} \bigcup_{\mathfrak{a} \in \Xi} A_{\tau'}^{\mathfrak{a}} \Big) = 0.$$

This implies the other inequality for the Hausdorff dimension.

The case $\tau = 2$ follows from the fact that $\tau_x \geq 2$ for every irrational number x (see proposition 2.3), while by the above argument the set $\{x : \tau_x > 2\}$ has vanishing Lebesgue measure. □

PROOF OF THEOREM 3.6. The set $\{x : \beta(x) = \delta\}$ is completely determined by theorems 3.2 and 3.3. Its Hausdorff dimension in the case of cuspidal f is immediate, while if f is non-cuspidal it follows from corollary 3.18. □

3.7. Examples

In the rest of this chapter we are going to apply the developed machinery to some interesting examples, namely Jacobi's theta function and newforms for $\Gamma_0(N)$. The included graphics have been plotted using *SageMath* [**84**], and the same software system has been used to compute the Fourier coefficients of newforms. The data points were calculated using simple C++ programs.

3.7.1. "Riemann's example". We are going to discuss some features of the graph of Riemann's example (**I**.9), plotted in figure **I**.2. The material in this section is not new: a similar but more detailed exposition is given by Duistermaat in [**24**]. Our analysis, however, is readily applicable to any other modular form.

Riemann's example φ satisfies $2\varphi(x) = \Im\theta_1(x)$. As we discussed in chapter 2, Jacobi's theta function θ is a modular form of weight $1/2$ for the theta group Γ_θ, consisting of all matrices in $\mathrm{SL}_2(\mathbb{Z})$ of the form $\left(\begin{smallmatrix}\text{odd} & \text{even}\\ \text{even} & \text{odd}\end{smallmatrix}\right)$ or $\left(\begin{smallmatrix}\text{even} & \text{odd}\\ \text{odd} & \text{even}\end{smallmatrix}\right)$. The Γ_θ-orbit of 0 corresponds to ∞ together will all the rationals p/q with either p even and q odd, or p odd and q even. All the remaining rationals (p/q with both p and q odd) constitute the Γ_θ-orbit of 1. The modular form θ is cuspidal at 1 but not at 0 and the associated multipliers μ_γ are always 8th roots of unity. All theses facts are proved assuming no background knowledge in Duistermaat's exposition [**24**], although they can also be deduced with some work from proposition 2.7.

A direct application of the regularity theorems suffices to recover Hardy's and Gerver's theorems (see §**I**.2) and determine the Hölder exponents of φ at every point. Its spectrum of singularities, first obtained by Jaffard in [**64**], also follows from theorem 3.6.

Jacobi's function θ is classically denoted ϑ_3, as it has two companions which are also modular forms of weight $1/2$ for conjugated groups of Γ_θ (*cf.* proposition 2.7):

$$\tilde\theta(z) = \vartheta_2(z) = \sum_{n\in\mathbb{Z}} e^{\left(n+\frac12\right)^2\pi iz} \quad \text{and} \quad \theta(z+1) = \vartheta_4(z) = \sum_{n\in\mathbb{Z}} (-1)^n e^{n^2\pi iz}.$$

The nomenclature $\tilde\theta$ is not standard but we employ it here as a convenient way to avoid problems with subscripts.

By proposition 2.7, given any matrix $\sigma \in \mathrm{SL}_2(\mathbb{Z})$ the modular form θ^σ is either a constant multiple of $\vartheta_2 = \tilde\theta$, $\vartheta_3 = \theta$ or $\vartheta_4(z) = \theta(z+1)$, the constant being an 8th root of unity. Since θ^σ is cuspidal at ∞ if and only if $\theta(\sigma(\infty)) = 0$, one concludes that:

$$\theta^\sigma(z) = \begin{cases} C\theta(z) \text{ or } C\theta(z+1) & \text{if } \sigma(\infty) \in [0], \\ C\tilde\theta(z) & \text{if } \sigma(\infty) \in [1]. \end{cases}$$

We now apply theorem 3.4 with $\alpha = 1$, $r = 1/2$, to study the behavior of $\varphi = \frac12\Im\theta_1$ in the neighborhood of a given rational point x_0. The resulting expansion

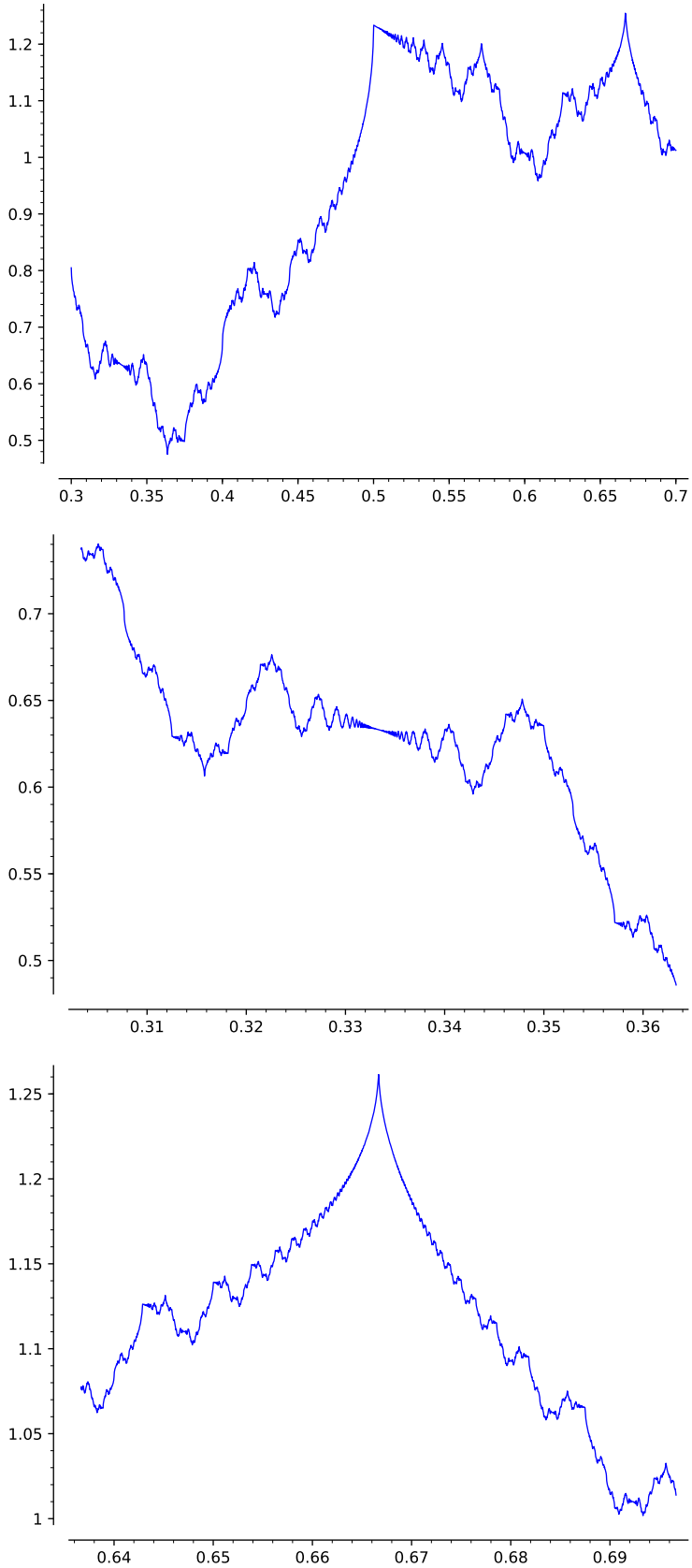


FIGURE 3.1. Detail of φ near 1/2, 1/3 and 2/3, respectively.

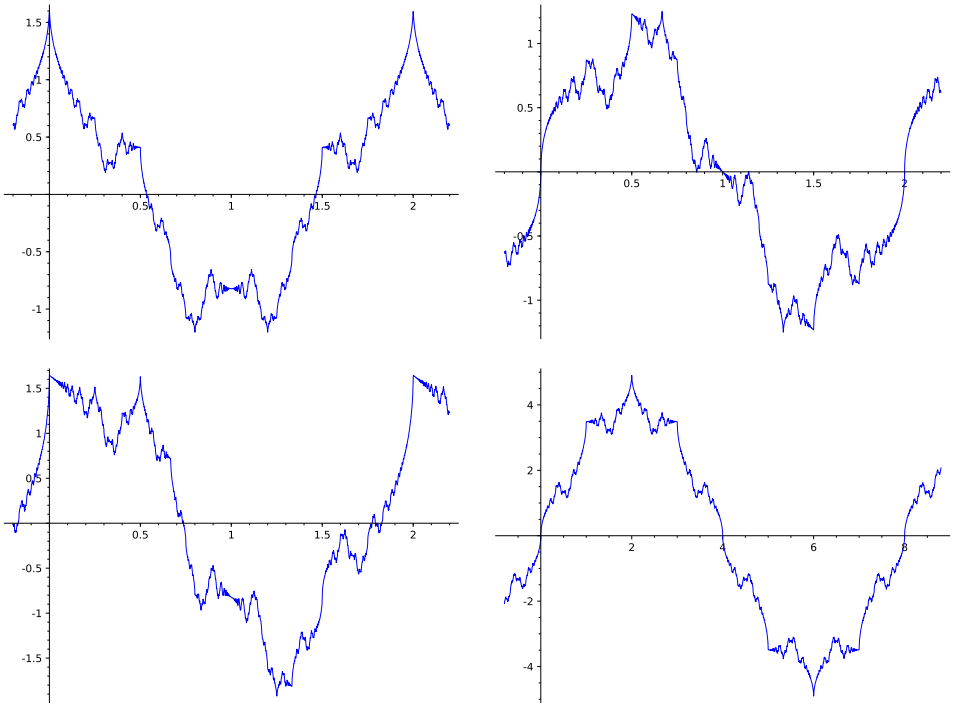


FIGURE 3.2. Graphs of $\Re\theta_1$ (top-left), $\Im\theta_1$ (top-right), $\Re\theta_1+\Im\theta_1$ (bottom-left) and $\Im\tilde{\theta}_1$ (bottom-right).

around x_0 is of the form:

$$\varphi(x) = \Im\left[C\sqrt{x-x_0}\right] + \Im\left[C'(x-x_0)^{3/2} f_1(\sigma^{-1}x+\tau)\right] + h(x).$$

The constant C is nonzero if and only if $x_0 \in [0]$, and in this case $f = \theta$. Otherwise $f = \tilde{\theta}$. The constant C' is always nonzero, and both constants have the argument of an 8th root of unity. Finally, τ is either 0 or 1.

Some deductions are immediate. The first one being that φ has singularities of square root type at every rational of the form odd/even or even/odd (either at one side or both sides of the rational). The second one is that at either side of any rational number φ mimics the graph of some periodic function, namely $\Im C' f_1$ if $x > x_0$ or $-\Re C' f_1$ if $x < x_0$. Note that as σ^{-1} has a simple pole at x_0, this pattern repeats indefinitely towards the rational, with its amplitude decreasing as a 3/2 power of the remaining distance and its frequency roughly proportional to $|x-x_0|^{-1}$. See figure 3.1 for some examples of this behavior, where some square root singularities are also clearly visible.

Since the argument of C' is an integer multiple of $\pi/4$ we also deduce that $\Im C' f_1$ (or $-\Re C' f_1$) is, up to a positive constant factor, either $\Re f_1$, $\Im f_1$ or $\Re f_1 + \Im f_1$, or the mirror image of one of these three functions, *i.e.*, the result of performing the change of variables $x \mapsto -x$ either in the domain, in the codomain or both. The situation is even simpler when $f = \tilde{\theta}$, as the functional equation $\tilde{\theta}(z+1) = \sqrt{i}\,\tilde{\theta}(z)$ implies that all these functions are then translates of each other and therefore we need only to consider $\Im\tilde{\theta}_1$. Hence the graph of $\Im C' f_1$ (or $-\Re C' f_1$) corresponds, up to symmetries, to one of the four genuinely distinct patterns that appear in figure 3.2. Note that in figure 3.1 all four patterns appear.

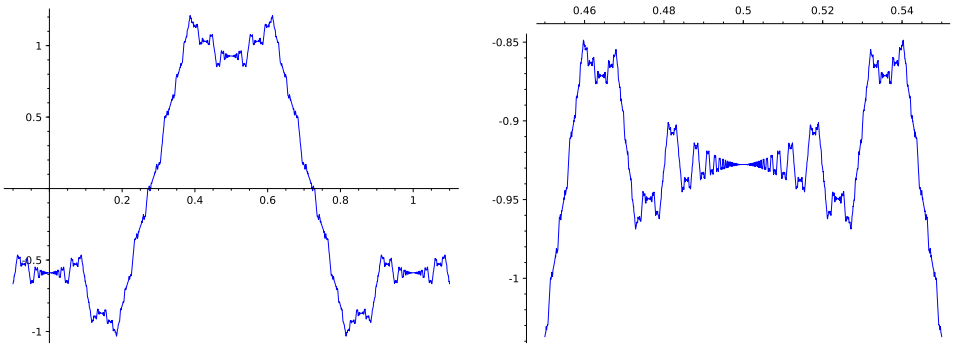


FIGURE 3.3. Left: Plot of $-\Re f_{9/5}$, where f is the newform on $\Gamma_0(14)$. Right: detail of $\Re f_{9/5}$ at $1/2$. This rational is not in $[\infty]$, but the matrix $\sigma = \left(\begin{smallmatrix} 7 & 3 \\ 14 & 7 \end{smallmatrix}\right)$ satisfies $\sigma(\infty) = 1/2$ and $f|_\sigma = -f$.

A different kind of self-similarities, modulo a $\mathcal{C}^{1,0}$ error term, may be found around fixed points of transformations lying in Γ_θ, as deduced from theorem 3.4 by letting x approach the fixed point of the transformation. Note that by theorem 1.4 this includes all quadratic surds. In this case the "zooming" factor given by the derivative of σx at the fixed point x_0 has magnitude different than 1 (as $j_\sigma(x_0)$ is irrational), and therefore the pattern repeats in a geometric progression towards x_0.

3.7.2. Cusp forms for $\Gamma_0(N)$. Fix an arbitrary integer $N \geq 1$ and let f be a cusp form of integer weight r for the group $\Gamma_0(N)$ and trivial multiplier system. Note that r must necessarily be an even integer. For any $\alpha > r/2$ the function f_α is well-defined and we may consider $g = \Re f_\alpha$ or $\Im f_\alpha$. Under these conditions by theorem 3.4 we have for every rational $x_0 \in [\infty]$,

$$g(x) = B|x - x_0|^{2\alpha - r} g(\sigma^{-1}x) + E(x) \tag{3.25}$$

for some $\sigma \in \mathrm{SL}_2(\mathbb{R})$ satisfying $\sigma^{-1}x_0 = \infty$ and the function E lying in the spaces $\mathcal{C}^{1,0}(\mathbb{R} \setminus \{x_0\})$ and $\mathcal{C}^{2\alpha - r + 1}(x_0)$. An interesting question is whether an approximate functional equation of the form (3.25), with B real and E with the same regularity, relating g with itself, exists for other rational numbers. Note this will happen for the rational x_0 as long as we are able to find some $\sigma \in \mathrm{SL}_2(\mathbb{R})$ satisfying $\sigma^{-1}x_0 = \infty$ and such that $f^\sigma = f|_\sigma$ equals Cf for a real constant C (and this is likely a necessary condition). In this section we provide sufficient conditions for this to hold and study some examples.

Some notation first. For every prime p we denote by $[n]_p$ the largest power of p dividing n, and for every divisor $Q \mid N$ satisfying $\gcd(Q, N/Q) = 1$ we define the matrix

$$\omega_Q := \begin{pmatrix} Qx & y \\ Nz & Qw \end{pmatrix}, \qquad x,\, y,\, z,\, w \in \mathbb{Z}, \quad \det \omega_Q = Q,$$

which is unique up to left and right multiplication by elements of $\Gamma_0(N)$. The matrices ω_Q are called Atkin-Lehner involutions and satisfy $Q^{-1}\omega_Q^2 \in \Gamma_0(N)$ and $\omega_Q \omega_{Q'} =$ some $\omega_{QQ'}$ whenever $\gcd(Q, Q') = 1$. For the sake of clarity we also set $\omega_p := \omega_{[N]_p}$ for each prime $p \mid N$. Finally for any integer $n > 0$ we consider the matrix

$$S_n := \begin{pmatrix} 1 & 1/n \\ 0 & 1 \end{pmatrix},$$

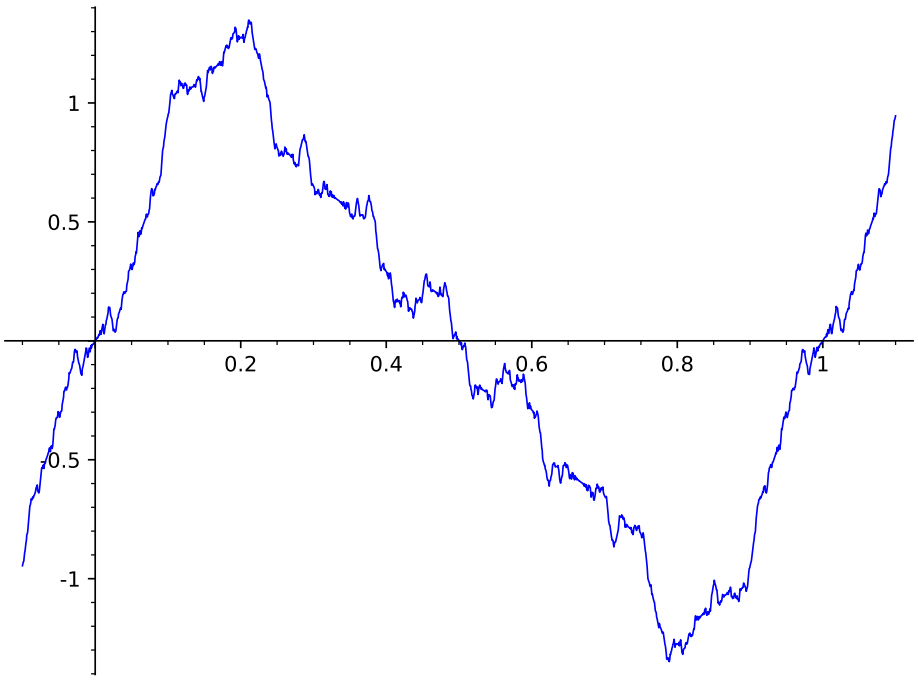


FIGURE 3.4. Plot of $\Im f_{7/4}$ where f is the newform on $\Gamma_0(45)$.

which corresponds to a translation by $1/n$.

Note that if we want $f|_\sigma = Cf$, then f must be a modular form for both $\Gamma_0(N)$ and $\sigma^{-1}\Gamma_0(N)\sigma$. Hence a good place to look for σ is in the normalizer of $\Gamma_0(N)$. A theorem of Atkin and Lehner stated without proof in [**2**] assures that when N is not divisible by 4 nor 9 this normalizer is generated by $\Gamma_0(N)$ and the Atkin-Lehner involutions ω_p for primes $p \mid N$. When N is divisible by 4 or by 9 one has to include some extra generators: S_2 if $[N]_2 = 4$ or 8, S_4 if $[N]_2 = 16$ or 32 and S_8 if $64 \mid N$; and S_3 if $9 \mid N$. Note that we are considering the normalizer of $\Gamma_0(N)$ as a group of linear fractional transformations, as otherwise one also needs to include any real multiple of the previous generators. This theorem also provides the structure of the quotient group of the normalizer of $\Gamma_0(N)$ over $\Gamma_0(N)$ itself (which we do not need), although this part seems to have some mistakes and a corrected version is proved by Bars in [**3**].

Asai observed in [**1**] that the Atkin-Lehner involutions act transitively on $\mathbb{Q}$ if and only if N is square-free. The following proposition is a generalization of this fact.

PROPOSITION 3.19. *The normalizer of $\Gamma_0(N)$ acts transitively on $\mathbb{Q}$ if and only if $N = 2^a 3^b N'$ for some $a < 8$, $b < 4$ and a square-free integer N' not divisible by 2 nor 3.*

PROOF. Assume first that N is of the prescribed form and let u/v be an arbitrary rational number, $\gcd(u, v) = 1$. It suffices to show that u/v is related modulo the normalizer to some u'/v' with $\gcd(u', v') = 1$ and $N \mid v'$, as these rationals comprise the orbit of ∞ modulo $\Gamma_0(N)$. We do this by stages, first relating it to a rational whose denominator is divisible by N', then adding 2^a and finally 3^b.

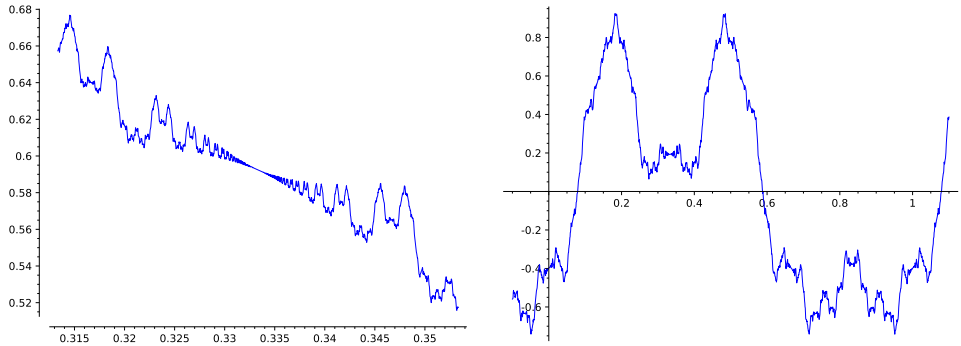


FIGURE 3.5. Left: Detail of $\Im f_{7/4}$ around $1/3$ where f is the newform on $\Gamma_0(45)$. Right: Graph of the imaginary part of the right hand side of (3.26).

Write $N' = p_1 \cdots p_n$ for distinct primes $p_1, \ldots, p_n$. We may assume upon reordering of the p_i that $p_1 \cdots p_m \mid v$ and $p_i \nmid v$ for $m < i \leq n$. Choosing $Q = 2^a 3^b p_{m+1} \cdots p_n$ we have

$$u'/v' = \omega_Q(u/v) = \frac{Qxu + yv}{N\left(zu + w\frac{v}{N/Q}\right)}.$$

The numerator of the right hand side is not divisible by any of the p_i as a consequence of the determinant condition imposed on ω_Q and therefore $N' \mid v'$.

Hence assume that from the beginning $N' \mid v$. This divisibility property is preserved by ω_2, S_2, S_4 and S_8. We show now we may find a related u'/v' with $2^a N' \mid v'$. Let $2^s = [v]_2$ and assume that $s < a$, since otherwise we are finished. It is easy to check that if $u'/v' = \omega_2(u/v)$ then $[v']_2 = 2^{a-s}$ if $2 \nmid u$ and $2^a \mid v'$ if $2 \mid u$. In the latter case we are finished, while in the former applying ω_2 if necessary we may assume $s \leq \lfloor a/2 \rfloor$. We now apply repeatedly S_2, S_4 or S_8 to arrive to a rational with $s = 0$, and the image of this rational by ω_2 satisfies $s \geq a$.

The same argument can now be applied *mutatis mutandis* to add the factor 3^b to the denominator. This finishes the proof of the direct implication.

To prove that the normalizer action is not transitive when N is not of the prescribed form it suffices to show a proper subset of $\mathbb{Q}$ invariant under this action. Suppose first that for some prime $p \neq 2, 3$ we have $p^2 \mid N$ and $p^c = [N]_p$. Then one such set is that of the rational numbers u/v with $[v]_p = p^s$ and $0 < s < c$. The invariance of this set follows from the following facts: the translations and the Atkin-Lehner involutions ω_Q with $p \nmid Q$ leave $[v]_p$ invariant, while $[v']_p = p^{c-s}$ for $u'/v' = \omega_Q(u/v)$ with $p \mid Q$.

The remaining cases are $2^8 \mid N$ or $3^4 \mid N$. If $2^8 \mid N$ then $a \geq 8$ and one such set is that of the rational numbers u/v with $[v]_2 = 2^{a/2}$ if a is even and $[v]_2 = 2^{\lfloor a/2 \rfloor}$ or $[v]_2 = 2^{\lfloor a/2 \rfloor + 1}$ if a is odd. An analogous set works when $3^4 \mid N$. □

If a cusp form f satisfies $f|_\sigma = C_\sigma f$, where C_σ is a real constant, for every σ lying in the normalizer of $\Gamma_0(N)$, then we may guarantee the approximate functional equation (3.25) to exist around every rational number in the orbit of ∞ modulo this normalizer, and hence around every rational if the action of the normalizer is transitive. Suppose now that f is a newform (see §2.9). Atkin and Lehner proved in [**2**] that $f|_{\omega_p} = \pm f$ for every prime $p \mid N$. In the same paper they also prove that

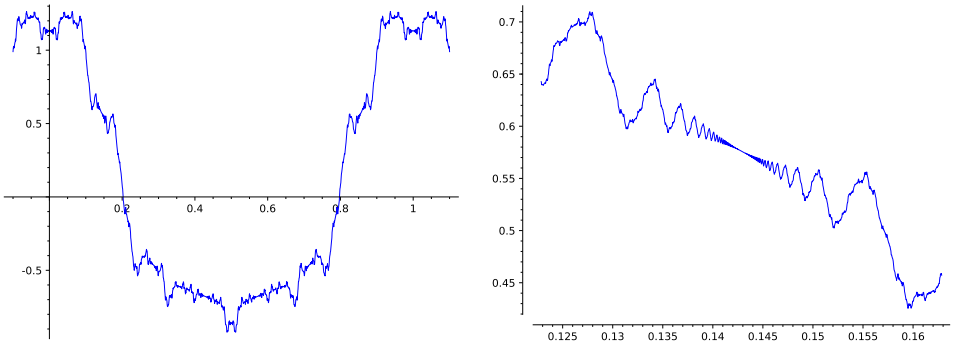


FIGURE 3.6. Left: Plot of $\Re f_{7/4}$ where f is the newform on $\Gamma_0(49)$. Right: Detail around $1/7$.

when $4 \mid N$ all the even coefficients of f vanish, and therefore $f|_{S_2} = -f$. If these transformations suffice to generate the normalizer, then the previous remarks apply.

When we have to include S_3, S_4 or S_8 to generate the normalizer, however, this breaks down, as it is not generally true that $f|_{S_n} = Cf$ for a real constant C. A workaround exists when the space of cuspidal forms has dimension 1. In this case, $f|_\eta$ is again a constant multiple of f for any η in the normalizer of $\Gamma_0(N)$, and therefore all these matrices commute under the action of the slash operator. As a consequence, $f|_\eta = f|_{\omega_Q S} = \pm f|_S$ for some $Q \mid N$ and some translation S. The matrix $\sigma = \eta S^{-1}$ now lies in the normalizer of $\Gamma_0(N)$ and satisfies $\sigma(\infty) = \eta(\infty)$ and $f|_\sigma = \pm f$. Therefore, if the normalizer acts transitively on $\mathbb{Q}$, so does the subgroup consisting of those matrices σ for which $f|_\sigma = \pm f$.

We conclude that the following are sufficient conditions to ensure that there is an approximate functional equation (3.25) around every rational number: (i) $N = 2^a N'$ with $a < 4$ and N' odd and square-free, or (ii) the space of cusp forms on $\Gamma_0(N)$ has dimension 1 and $N = 2^a 3^b N'$ with $a < 8$, $b < 4$ and N' square-free and not divisible by 2 nor 3.

To end the section we give some examples of modular forms for which an equation like (3.25), relating g to itself, is unlikely to exist around some rational numbers. These are of weight 2 and therefore associated to modular abelian varieties over $\mathbb{Q}$. By direct examination of the table of newforms found at [**76**] we see that the lowest value of N for which neither of the previous conditions is satisfied is $N = 45$, as the associated space of cusp forms happens to be of dimension 3, containing an oldclass generated by the newform on $\Gamma_0(15)$. Denote by f the newform on $\Gamma_0(45)$ and by h the one on $\Gamma_0(15)$. These are associated to the isogeny classes of the elliptic curves

$$y^2 + xy = x^3 - x^2 - 5 \quad \text{and} \quad y^2 + xy + y = x^3 + x^2,$$

respectively. The matrix $\sigma = S_3\omega_{45}$, where ω_{45} is the Atkin-Lehner involution determined by $x = w = 0$, $y = 1$ and $z = -1$, lies in the normalizer of $\Gamma_0(45)$ and sends ∞ to $1/3$. The function $f|_\sigma$ is therefore again a modular form for $\Gamma_0(45)$, and in fact it has the following decomposition:

$$f|_\sigma(z) = \frac{1}{2} f(z) - i\frac{1}{2\sqrt{3}} h(z) - i\frac{3\sqrt{3}}{2} h(3z).$$

To obtain the coefficients one first decomposes $f|_{S_3}$ by directly comparing coefficients, and then applies $|_{\omega_{45}}$. The Atkin-Lehner eigenvalues are tabulated in [**76**],

and the action of this operator on oldforms is described by lemma 26 of [**2**]. As an immediate consequence

$$f_\alpha^\sigma(x) = \frac{1}{2} f_\alpha(x) - \frac{i}{2\sqrt{3}} h_\alpha(x) - \frac{i}{2 \cdot 3^{\alpha-3/2}} h_\alpha(3x). \tag{3.26}$$

In figure 3.4 we have plotted $g = \Im f_{7/4}$, while in figure 3.5 the reader can compare the imaginary part of the right hand side of (3.26) for $\alpha = 7/4$ with aspect of the graph of g near $\sigma(\infty) = 1/3$.

The lowest value of N for which the normalizer is not transitive on $\mathbb{Q}$ and there is some nonzero newform is $N = 49$. This newform is associated to the isogeny class of the curve

$$y^2 + xy = x^3 - x^2 - 2x - 1.$$

The cusp $1/7$ is not related to ∞, not even by the normalizer, and in figure 3.6 the reader can appreciate how for $g = \Re f_{7/4}$ the aspect of the repeating pattern around $1/7$ and that of the global graph seem to differ, making it unlikely for a self-similarity relation like (3.25) to hold.

CHAPTER 4

Lattice point counting problems

In this chapter we provide a general framework for this family of problems, of which particular cases will be discussed in chapters 5 and 6, together with very general tools to address them.

4.1. Definitions and conjectures

Let $K \subset \mathbb{R}^d$ be a compact body with non-empty interior, and for every $R > 1$ denote by $\mathcal{N}_K(R)$ (or $\mathcal{N}(R)$ if there is no possible confusion) the number of points in $\mathbb{Z}^d$ lying in K after being dilated by the factor R, *i.e*,

$$\mathcal{N}(R) = \#\{\vec{n} \in \mathbb{Z}^d \,:\, \vec{n}/R \in K\}.$$

For convenience we will also use the notation $RK = \{\vec{x} \,:\, \vec{x}/R \in K\}$, so that $\mathcal{N}(R) = \#\mathbb{Z}^d \cap RK$. As described in the introduction, the lattice point counting problem associated to K consists in estimating the error term

$$\mathcal{E}(R) = \mathcal{N}(R) - |K|R^d,$$

where $|K|$ stands for the d-dimensional volume of K. Sometimes, when specified, we will replace in these definition either the way we count the points in $\mathcal{N}(R)$ or the main term $|K|R^d$ in $\mathcal{E}(R)$ with appropriate versions for the region at hand. In any case, we are interested in the optimal exponent

$$\alpha_K = \inf\{\alpha > 0 \,:\, \mathcal{E}(R) = O(R^\alpha)\}.$$

Under mild hypotheses, Lipschitz boundary for example, the argument by Gauss sketched in §**I**.3 shows $\alpha_k \leq d-1$, and the d-dimensional unit cube is an example where this is sharp. When there is curvature, however, one can usually do better. In this regard, we say that K is a *smooth convex body* if its boundary is a smooth $(d-1)$-dimensional submanifold of $\mathbb{R}^d$ whose Gaussian curvature is positive everywhere. The following table summarizes the best known upper bounds for the exponent α_K for smooth convex bodies and for the particular family of balls, and the conjectured value for both cases:

d	smooth convex body	d-dimensional ball	conjecture
2	$\alpha_K \leq 131/208$ Huxley [**59**]	$\alpha_K \leq 517/824$ Bourgain, Watt [**11**]	1/2
3	$\alpha_K \leq 231/158$ Guo [**39**]	$\alpha_K \leq 21/16$ Heath-Brown [**47**]	1
≥ 4	$\alpha_K \leq d-2+r(d)$ Guo [**39**]	$\alpha_K = d-2$	$d-2$

In the bottom-left entry, $r(d) = (d^2+3d+8)/(d^3+d^2+5d+4)$.

Some comments on these results. The bound for the exponent for bidimensional smooth convex bodies was obtained by Huxley, and until very recently it was also the best known upper bound for the exponent of the Gauss' circle problem (unit disk). Bourgain and Watt used decoupling to improve it to 517/824 for both the Gauss

circle problem and the Dirichlet divisor problem.[1] In principle the same techniques should yield the same exponent, or at least an improvement, over Huxley's result.

In three or more dimensions the best known result for smooth convex bodies is due to Guo, who used a bidimensional version of the van der Corput method. For the three-dimensional ball, the exponent 21/16 was obtained by Heath-Brown, building upon previous ideas of Chamizo and Iwaniec [**17**]. The same technique was also applied successfully by Chamizo, Cristobal and Ubis [**16**] to rational ellipsoids in three dimensions.

The result for balls in four or more dimensions is classic, and we provide a proof below based on Jacobi's four square theorem for the case where the ball is centered at the origin. The same exponent also applies to rational ellipsoids. The heuristic here is that the characteristic function of the set of the squares $\{n^2\}$ is a very arithmetic function, but as a one convolves it with itself it gains regularity. Hence r_2 is fairly arithmetic, r_3 shows regularity if one stays away from some "bad" values of the argument, $r_4(n)$ oscillates slightly between $n/\log\log n$ and $n\log\log n$ and $r_k(n) \asymp n^{k/2-1}$ for $k \geq 5$ (*cf.* corollary 11.3 of [**61**]). Since the sum $\sum_{n\leq R^2} r_k(n)$ does some further regularization, the conjectured exponent is obtained for dimension 4 too. The inequality $\alpha_K \geq d-2$ also follows from these asymptotics for $r_k(n)$, as the error term $\mathcal{E}(R)$ has jump discontinuities of size the number of points with integer coordinates lying on the boundary of K, and therefore it is an Ω-function of this quantity. In particular, $\mathcal{E}(R) = \Omega(R^{d-2})$ for $d \geq 3$ (for $d = 3$ see [**26**]). Similar arguments work if r_k is replaced by r_Q for an arbitrary rational quadratic form Q in k variables.

When $d \geq 5$ then the error term for both balls and rational ellipsoids satisfies the upper bound $\mathcal{E}(R) = O(R^{d-2})$, and hence the ϵ may be dropped. This is also classical. For a proof of this result we refer the reader to Fricker's book [**32**] (Satz 1 of §21).

For irrational ellipsoids much less is known. The inequality $\alpha_{\mathcal{K}} \leq d-2$ was finally achieved by Bentkus and Gotze in [**5**] for $d \geq 9$ and later Gotze extended the result in [**37**] to $d \geq 5$. Surprisingly, the error term is, in contrast with the rational case, $\mathcal{E}(R) = o(R^{d-2})$, and this led to a proof of a conjecture by Davenport and Lewis stating that the gaps of the image of $\mathbb{Z}^d$ under irrational quadratic forms tend to zero as one gets further away from the origin [**6**].

The conjectured exponent 1/2 for the circle problem comes from a lower bound for α_K proved independently by Hardy and Landau [**41, 73**]. In other cases the conjectured exponents are folklore.

We will comment some more results on lattice point counting problems in the subsequent chapters, and prove some new ones. The interested reader may consult the survey [**60**] for further information on this topic.

We sketch now the proof of $\alpha_K \leq d-2$ when K is the unit d-dimensional ball and $d \geq 4$. The first observation is that it suffices to prove $\mathcal{N}(R) = CR^4 + O(R^2\log R)$ for $d = 4$ and some constant C. Indeed, this implies $\mathcal{N}(R) = C_d R^d + O(R^{d-2}\log R)$ for every $d \geq 5$ by slicing the d-dimensional ball or radius R into parallel $(d-1)$-dimensional balls, applying the estimation to each of them and summing up all the resulting $(d-1)$-dimensional volumes with help of the Euler-Maclaurin formula. This bound for $\mathcal{E}(R)$ is not sharp, and in fact $\mathcal{E}(R) = O(R^{d-2})$ for $d \geq 5$, but it

[1] In the latter case, $\mathcal{E}(N) = \sum_{n\leq N} \sigma_0(n) - N\log N - N(2\gamma-1)$ where γ is the Euler-Mascheroni constant, and $R = N^{1/2}$.

suffices to obtain the right exponent. Note also that C_d must equal the volume of the d-dimensional unit ball by Gauss' argument.

Hence assume $d = 4$. Then by Jacobi's theorem $\mathcal{N}(R) = \sum_{n\le R^2} r_4(n)$, and $r_4(n) = 8\sum_{d|n} d$ if n is odd and $r_4(n) = 24\sum_{d|n,\, d\,\mathrm{odd}} d$ if n is even. The idea is to use Dirichlet's hyperbola method to estimate this double sum with a good error term. We will only do this for the first sum, as an analogous computation gives the asymptotics for the second one. Put $S = \sum_{n\le N,\, n\,\mathrm{odd}} r_4(n)$ and write

$$\begin{aligned} S &= \sum_{\substack{n\le N\\ n\text{ odd}}} \sum_{\substack{d|n\\ d\le\sqrt{n}}} d \;+\; \sum_{\substack{n\le N\\ n\text{ odd}}} \sum_{\substack{d|n\\ d\le\sqrt{n}}} \frac{n}{d} \;-\; \sum_{\substack{n\le N\\ n\text{ odd square}}} \sqrt{n} \\ &= \sum_{\substack{d\le\sqrt{N}\\ d\text{ odd}}} d \sum_{\substack{d\le d_1\le N/d\\ d_1\text{ odd}}} 1 \;+\; \sum_{\substack{d\le\sqrt{N}\\ d\text{ odd}}} \sum_{\substack{d\le d_1\le N/d\\ d_1\text{ odd}}} d_1 \;+\; O(N) \\ &= \sum_{\substack{d\le\sqrt{N}\\ d\text{ odd}}} \left(\frac{N}{2} - \frac{3d^2}{4} + \frac{N^2}{4d^2} + O(N/d + d) \right) \;+\; O(N). \end{aligned}$$

Note now that $\sum_{d\le\sqrt{N},\, d\,\mathrm{odd}} d^{-2} = C - \sum_{d>\sqrt{N},\, d\,\mathrm{odd}} d^{-2}$, and by the Euler-Maclaurin formula this latter sum is $1/(2\sqrt{N}) + O(1/N)$. Hence $S = CN^2/4 + O(N\log N)$.

4.2. The exponential sum

Most results in lattice point counting theory are obtained by first translating the problem to that of bounding an exponential sum. To do this the characteristic function of the dilated body is smoothed by convolving by a mollifier and then Poisson summation is applied. This is the approach taken in the introduction to give a proof of Sierpiński's result, where we used that the Fourier transform of the characteristic function of the unit disk has a explicit expression for which good asymptotics hold. In general we cannot hope to be able to compute the Fourier transform of the characteristic function of K, but if K is assumed to be a smooth convex body then it is possible to obtain good asymptotics nevertheless. This was first done by Hlawka in [**51**], with the error term later improved by Herz [**50**]. We need the latter result. Although the proof provided by Herz is rather convoluted, the interested reader can find a much more down to earth approach in chapter 7 of Hörmander's book [**53**] (corollary 7.7.15). The result states that whenever $K \subset \mathbb{R}^d$ is a smooth convex body and χ its characteristic function,

$$\hat{\chi}(\vec{\xi}) = \frac{e\big(g(-\vec{\xi}) - (d-1)/8\big)}{2\pi i\|\vec{\xi}\|^{(d+1)/2}\sqrt{\kappa(-\vec{\xi})}} - \frac{e\big(-g(\vec{\xi}) + (d-1)/8\big)}{2\pi i\|\vec{\xi}\|^{(d+1)/2}\sqrt{\kappa(\vec{\xi})}} + O\left(\frac{1}{\|\vec{\xi}\|^{(d+3)/2}}\right), \tag{4.1}$$

where $g(\vec{\xi}) = \sup\{\vec{x}\cdot\vec{\xi} \,:\, \vec{x}\in K\}$ and $\kappa(\vec{\xi})$ stands for the Gaussian curvature at the point whose unit outer normal is $\vec{\xi}/\|\vec{\xi}\|$. The proof essentially consists in an application of the *stationary phase principle*. This principle states that the main contribution to an integral of the form $\int f(\vec{x})e\big(t\phi(\vec{x})\big)\,d\vec{x}$ as $t\to\infty$ comes principally from neighborhoods of the zeros of $\nabla\phi(\vec{x})$, *i.e.* where the phase function ϕ becomes stationary, as for the rest of points $t\phi$ changes rapidly and for reasonably good functions f the integral has a fair amount of cancellation. Now, $\hat{\chi}(\vec{\xi}) = \int_K e(-\vec{x}\cdot\vec{\xi})\,d\vec{x}$, and this integral practically vanishes except in a thin neighborhood around the boundary ∂K of K. There the phase becomes stationary at the zeros of $\nabla(\vec{x}\cdot\vec{\xi})|_{\partial K}$, *i.e.* at the points where the suprema defining $g(\vec{\xi})$ and $g(-\vec{\xi})$ are attained. By

geometric considerations these are points where the unit outer normal equals $\pm\vec{\xi}/\|\vec{\xi}\|$. Hence $\hat{\chi}(\vec{\xi}) \approx \int_{\partial K} e\big(-g(\vec{\xi}) + H^+(\vec{x})\big)\,d\vec{x} + \int_{\partial K} e\big(g(-\vec{\xi}) + H^-(\vec{x})\big)\,d\vec{x}$ where $H^\pm$ are the Hessians at the points where the unit outer normal equals $\pm\vec{\xi}/\|\vec{\xi}\|$, and whose determinant is the Gaussian curvature at those points. Estimating these integrals one arrives to (4.1).

We are going to carry out the argument sketched above to relate the error term in the lattice point count problem $\mathcal{E}(R)$ to the corresponding exponential sum explicitly in the case $d = 3$, as we only need this case. This is contained in the following proposition. Of course, an analogous result may be obtained from (4.1) with little effort for any $d \geq 2$.

PROPOSITION 4.1. *Let $K \subset \mathbb{R}^3$ be a smooth convex body. Let η be a smooth even function with support inside $[-1,1]$ and satisfying $\eta(0) = 1$ and that the Fourier transform of $\eta(\|\vec{x}\|)$ is a non-negative function. Fix $\epsilon > 0$ and $0 < c < 2$. Then for any $R > 2$ there exists $R' \in (R-1, R+1)$ satisfying*

$$\mathcal{E}(R) = -\frac{R'}{\pi} \sum_{\vec{0} \neq \vec{n} \in \mathbb{Z}^3} \eta(\delta\|\vec{n}\|) \frac{\cos\big(2\pi R' g(\vec{n})\big)}{\|\vec{n}\|^2 \sqrt{\kappa(\vec{n})}} + O(R^{2+\epsilon}\delta) \tag{4.2}$$

for $\delta = R^{-c}$ and g and κ as before.

This kind of results are usually regarded in the literature as *truncated Hardy-Voronoï formulas*, as Voronoï [**92**] was the first to prove an explicit formula for the sum $\sum_{n\geq 0} \sigma_0(n)\eta(n)$ where η is a smooth function of fast decay. An analogous formula for $\sum_{n\geq 0} r_2(n)\eta(n)$ was also suggested by Voronoï [**93**], and later rigorously proved independently by Sierpiński [**89**] and Hardy [**41**]. These formulas may be truncated with an error term to obtain a rather similar expression to (4.2) for Dirichlet's divisor and Gauss' circle problems.

To apply proposition 4.1 note that we can construct a function η satisfying all the hypotheses by picking a real nonzero even smooth function η_1 supported in $[-1/2, 1/2]$ and then choosing $\eta(x) = C\eta_2 * \eta_2(x,0,0)$ where $\eta_2(\vec{x}) = \eta_1(\|\vec{x}\|)$ and $C > 0$ is an appropriate constant. As η_2 is radial, so is $\eta_2 * \eta_2$ and $\eta(\|\vec{x}\|) = C\eta_2 * \eta_2(\vec{x})$. Hence the Fourier transform of this function equals $C\hat{\eta_2}^2$, which is non-negative as $\hat{\eta_2}$ is real because η_2 is an even real function. Despite this very particular construction, we have a fair amount of freedom to choose η. It is interesting that since neither $\mathcal{E}(R)$ nor the order of magnitude of the error term depend on η, the exponential sum must have the same amount of cancellation independently of how we are truncating it.

To have an idea of how powerful this result is we may bound the sum on the right hand side of (4.2) term by term, disregarding all cancellation, to obtain $\mathcal{E}(R) \ll R^{1+c} + R^{2-c+\epsilon}$. Choosing $c = 1/2$ we have $\mathcal{E}(R) \ll R^{3/2+\epsilon}$ for all $\epsilon > 0$, *i.e.* $\alpha_K \leq 3/2$. This is the analogue of Sierpiński's result for the circle. The very same argument carried on for an arbitrary number of dimensions $d \geq 2$ shows $\alpha_K \leq d(d-1)/(d+1)$, a result first obtained by Hlawka in [**51**].

To gain some intuition when there is cancellation it is better to consider first what happens with an unidimensional sum $S = \sum_{n\leq N} n^\alpha e\big(\phi(n)\big)$ for some α. Summing by parts, $S \ll \sum_{n\leq N-1} |S_n| n^{\alpha-1} + |S_N| n^\alpha$, where $S_N = \sum_{n\leq N} e\big(\phi(n)\big)$. These exponential sums, if the values of $\phi(n)$ modulo 1 are uncorrelated, should be expected to have square root cancellation, *i.e.* to be of size $N^{1/2}$. If the values of $\phi(n)$ are only "slightly" correlated, one should expect the size of the exponential sum to

increase as a power of N between the square root bound and the trivial bound N. Let us say $|S_N| \ll N^{1-s}$. Substituting above, $S \ll N^{1+\alpha-s} + \log N$, and hence up to a logarithm (which may not appear) we have gained N^{-s} over the trivial bound $N^{1+\alpha}$ which is obtained by estimating the original sum termwise. The same heuristics apply for sums for the form $\sum_{n\leq N} f(n)e(\phi(n))$ where f is a reasonably good function.

Back to the formula (4.2), suppose after summing by parts in three variables we have a power savings of order N^{-s} in the exponential sum. Since the sum is of "length" R^{3c}, we should expect a bound $\mathcal{E}(R) \ll R^{1+c-3cs} + R^{2-c+\epsilon}$ and taking $c = 1/(2-3s)$ (for $s \leq 1/2$) we obtain $\alpha_K \leq 1 + (1-3s)/2$. The conjecture would therefore be obtained for $s = 1/3$. This might seem feasible, being far away from square root cancellation, but the current methods for handling d-dimensional exponential sums are very poor. Not only that, but also for these exponential sums we cannot expect nothing close to square root cancellation to hold, and $s = 1/3$ seems to be at the boundary of what is true as shown, for example, by the known Ω-results for the sphere. In fact, it is a better idea to think of the sum as a triple sum, in each of the variables n_1, n_2, n_3, each one of length R^c. Then the conjecture corresponds to having square-root cancellation in two of the three sums, as then we would have the bound $R^{c+c/2+c/2} = \left(R^{3c}\right)^{1-s}$ for $s = 1/3$. The third sum would then provide no additional cancellation. The same heuristics carried on in dimension $d \geq 2$ show that for $d = 2$ the conjecture corresponds to having square-root cancellation in only one of the two iterated sums, and for $d \geq 3$ we expect to have square-root cancellation in two of the d iterated sums.

Proof of proposition 4.1. (Adapted from proposition 2.1 of [**15**]) We prove the result assuming first that K contains the origin in its interior. Also, without loss of generality, we may assume ϵ is arbitrarily small, in particular $0 < \epsilon < 1$. Let ϕ be the Fourier transform of the function $\eta(\|\cdot\|)$, and write $\phi_\delta(\vec{\xi}) = \delta^{-3}\phi(\vec{\xi}/\delta)$, the Fourier transform of $\eta(\delta\|\cdot\|)$. Since $\int \phi = \eta(0) = 1$ and ϕ is of fast decay, for every $k \geq 1$ we have

$$\int\limits_{\|\vec{t}\|\leq\delta^{1-\epsilon}} \phi_\delta(\vec{t})\,d\vec{t} = 1 + O(\delta^k) \qquad \text{and} \qquad \int\limits_{\|\vec{t}\|\geq\delta^{1-\epsilon}} \phi_\delta(\vec{t})\,d\vec{t} = O(\delta^k)$$

as $\delta \to 0^+$. This is, almost all the mass is concentrated in the ball of radius $\delta^{1-\epsilon}$.

As K is convex with smooth boundary, there is some constant $C > 0$ such that for r small enough, any ball of radius r with the center inside K lies entirely inside $(1+Cr)K$ and any ball of radius r whose center is not in K lies entirely outside $(1-Cr)K$ (see figure 4.1). Taking $r = R^{-1}\delta^{1-\epsilon}$ and dilating by R we have

$$(\phi_\delta * \chi_{R_1})(\vec{x}) \leq \chi_R(\vec{x}) + O(\delta^k) \qquad \text{and} \qquad (\phi_\delta * \chi_{R_2})(\vec{x}) \geq \chi_R(\vec{x}) + O(\delta^k),$$

where χ_R stands for the characteristic function of RK, $R_1 = R - C\delta^{1-\epsilon}$ and $R_2 = R + C\delta^{1-\epsilon}$. This is the step where it is crucial that $\phi_\delta \geq 0$.

Hence, by the continuity of $\phi_\delta * \chi_R$ in R, there exists some R' such that $|R - R'| \leq C\delta^{1-\epsilon}$ and

$$\sum_{\vec{n}\in\mathbb{Z}^3} (\phi_\delta * \chi_{R'})(\vec{n}) = \mathcal{N}(R) + O(R^3\delta^k).$$

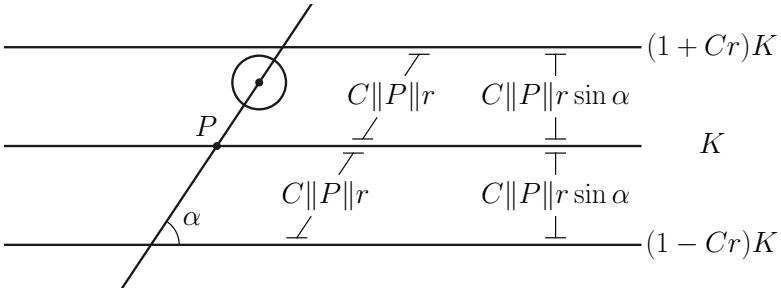


FIGURE 4.1. A ball of radius r (with unnamed center) lies outside K. Let P be the intersection point of the segment joining the center of the ball with the origin and the boundary of K. The tangent planes to K at P, to $(1+Cr)K$ at $(1+Cr)P$ and to $(1-Cr)K$ at $(1-Cr)P$ are parallel, drawn as horizontal lines in the picture, at a distance $C\|P\|r\sin\alpha$ to each other. As the actual boundary of $(1-Cr)K$ deviates from the tangent plane very little for small r, it suffices to choose C satisfying $C\|P\|\sin\alpha > 1$ to ensure that the ball cannot intersect $(1-Cr)K$.

In particular for δ small enough, $R-1 < R' < R+1$. Apply now Poisson summation to the sum on the left,

$$\mathcal{E}(R) = (R')^3|K| - R^3|K| + \sum_{\vec{0}\neq\vec{n}\in\mathbb{Z}^3} \eta(\delta\|\vec{n}\|)\hat{\chi}_{R'}(\vec{n}) + O(R^3\delta^k).$$

Substituting $\hat{\chi}_{R'}(\vec{\xi}) = (R')^3\hat{\chi}(R'\vec{\xi})$ in (4.1) above we obtain the estimation

$$\hat{\chi}_{R'}(\vec{n}) + \hat{\chi}_{R'}(-\vec{n}) = -\frac{R'}{\pi\|\vec{n}\|^2}\left(\frac{\cos\left(2\pi R' g(\vec{n})\right)}{\sqrt{\kappa(\vec{n})}} + \frac{\cos\left(2\pi R' g(-\vec{n})\right)}{\sqrt{\kappa(-\vec{n})}}\right) + O\left(\frac{1}{\|\vec{n}\|^3}\right).$$

Hence

$$\mathcal{E}(R) = -\frac{R'}{\pi}\sum_{\vec{0}\neq\vec{n}\in\mathbb{Z}^3} \eta(\delta\|\vec{n}\|)\frac{\cos\left(2\pi R' g(\vec{n})\right)}{\|\vec{n}\|^2\sqrt{\kappa(\vec{n})}} + O(R^2\delta^{1-\epsilon} + R^3\delta^k + \log\delta).$$

Substituting $\delta = R^{-c}$, renaming ϵ and choosing k big enough, the error term is $O(R^{2+\epsilon}\delta)$.

Suppose now that the origin does not lie in the interior of K. The number of points with integer coordinates inside RK does not vary if we translate K by a multiple amount of $1/R$ in any direction, and hence for all purposes we may replace K with a translation K' whose interior contains the origin, which is always possible for R big enough. The Fourier transform of RK also coincides with that of RK', leaving the same right hand side in (4.2). □

4.3. Vaaler-Beurling polynomials

An essential ingredient of the proof of proposition 4.1 was Poisson summation in all the variables. In chapter 5 however we will find a situation where it is convenient to do Poisson summation only in one of the variables to arrive to an exponential sum. This usually leads to weaker results than doing Poisson summation in every variable, as the resulting exponential sum is harder to manage. However the special geometry of the problem we will be concerned with, the paraboloids (*cf.* **I**.3), results in the exponential sum being as difficult to bound as the one obtained by full Poisson summation, and the lack of regularity of the boundary would require an *ad hoc* proof

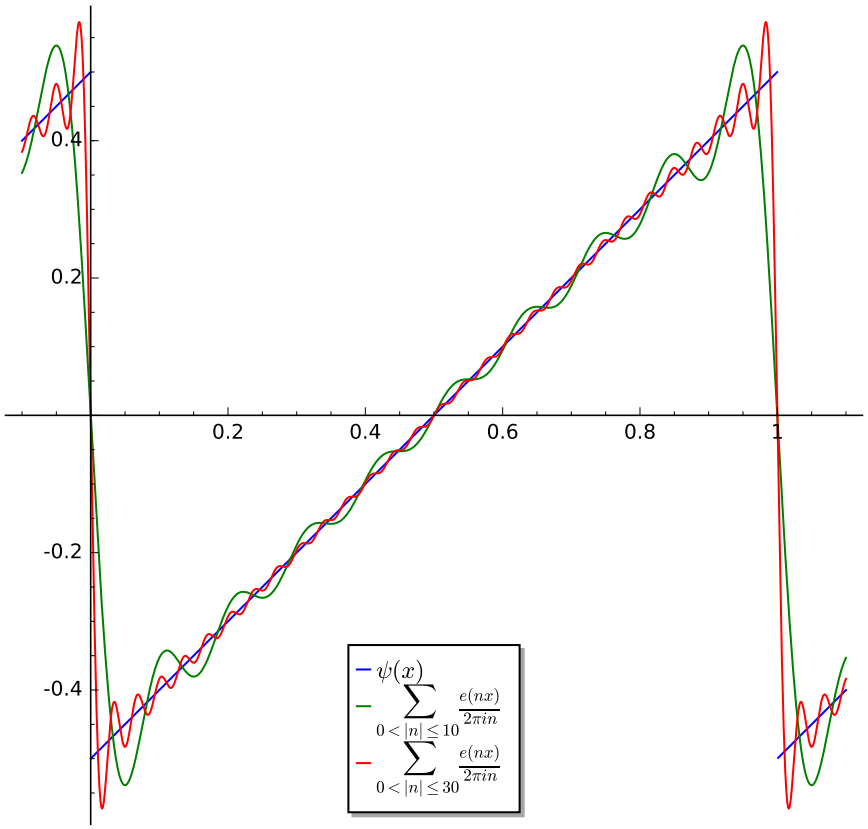


FIGURE 4.2. The saw-tooth function ψ and its approximation by two different truncated Fourier polynomials of degrees 10 and 30. The Gibbs phenomenon is the constant overshooting of the Fourier polynomials that can be seen close to the integer points.

of the asymptotics (4.1) for $\hat{\chi}$ in this case if we want to apply proposition 4.1 directly. More about this will be discussed in the chapter itself.

Here we present a "simple" way to do Poisson summation in one variable to transform a lattice point counting problem into an exponential sum. Consider the usual Poisson summation, $\sum_n f(n) = \sum_n \hat{f}(n)$. In the language of distributions this can also be written

$$\sum_{n\in\mathbb{Z}} \delta(x-n) = \sum_{n\in\mathbb{Z}} e(nx), \tag{4.3}$$

where δ is the usual *Dirac delta*: a "function" having the value 0 everywhere except at the origin, where it has the value ∞, and integrates 1. Indeed, multiplying by f, integrating in $\mathbb{R}$ and interchanging integration and summation be obtain the usual Poisson summation formula, and in fact by truncating these sums with an appropriate error term and taking the limit this leads to a different proof of the same result (note the truncated exponential sums are just the usual Dirichlet kernel). If we pass in (4.3) the term corresponding to $n = 0$ from the right hand side to the left hand side and formally integrate we obtain

$$-\psi(x) = \sum_{n\neq 0} \frac{e(nx)}{2\pi i n} \qquad \text{where} \qquad \psi(x) = x - \lfloor x \rfloor - 1/2. \tag{4.4}$$

This is the usual Fourier expansion for the *saw-tooth function* ψ, which actually converges to the right value for any $x \notin \mathbb{Z}$. This identity is also equivalent to Poisson summation, as can be shown by applying the Euler-Maclaurin formula to $\sum_n f(n)$ and substituting in the error term $\int f'(x)\psi(x)\,dx$ the Fourier expansion above. If we can interchange summation and integration, the resulting terms $(2\pi in)^{-1}\int f'(x)e(nx)$ can be integrated by parts back to $\hat{f}(n)$.

Suppose now that $K \subset \mathbb{R}^2$ is given by $|y| \leq f(|x|)$ for some function f strictly decreasing in $[0, x_0]$ with $f(x_0) = 0$, and denote by g the inverse function.[2] We are going to formally apply Poisson summation in the second variable of the sum $\sum_{\vec{n}} \chi_K * \eta(\vec{n})$ to see at which the exponential sum we would arrive if we follow the conventional route. After Poisson summation and subtracting the main term,

$$\mathcal{E} \approx \sum_{0\neq m\in\mathbb{Z}} \int \sum_{n\in\mathbb{Z}} \chi_K * \eta(n, y)e(-my)\,dy.$$

Expanding the definition of convolution and applying Fubini a couple of times,

$$\mathcal{E} \approx \sum_{0\neq m\in\mathbb{Z}} \int \hat{\eta}_s(m) \int G(s, t)e(-mt)\,dt\,ds$$

where $\eta_x(y) = \eta(x, y)$ and $G(s, t) = \sum_n \chi_K(n - s, t)$. It is not hard to see that $G(s, t)$ may also be written $\lfloor g(t) + \{s\}\rfloor + \lfloor g(t) - \{s\}\rfloor + 1$. Performing the change of variables $t = f(x)$ and integrating by parts, since $\frac{\partial}{\partial t}\lfloor t\rfloor = \sum_n \delta(n - t)$ we have

$$\begin{aligned}\mathcal{E} \approx &- \int \sum_{0\neq m\in\mathbb{Z}} \sum_{|n+\{s\}|\leq x_0} \hat{\eta}_s(m)\frac{e\big(-mf(n + \{s\})\big)}{2\pi im}\,ds \\ &- \int \sum_{0\neq m\in\mathbb{Z}} \sum_{|n-\{s\}|\leq x_0} \hat{\eta}_s(m)\frac{e\big(-mf(n - \{s\})\big)}{2\pi im}\,ds\end{aligned}$$

up to some boundary terms which are hopefully small.

Now let us do something different. Note $\mathcal{N} = 2\sum_n \lfloor f(n)\rfloor + 2\lfloor x_0\rfloor$ and substitute $\lfloor x\rfloor = x - 1/2 - \psi(x)$. Since $\sum_n f(n)$ can be sharply estimated via de Euler-Maclaurin formula, we also have

$$\mathcal{E} \approx -2\sum_{|n|\leq x_0} \psi(f(n)) \approx -2\sum_{0\neq m\in\mathbb{Z}} \sum_{|n|\leq x_0} \frac{e\big(mf(n)\big)}{2\pi im},$$

where we have substituted (4.4). The second approximation symbol is there because we do not know *a priori* how often $f(n)$ is an integer. Note this error term is similar to the one we had obtained via Poisson summation, except for the lack of the outer average in s and the mollifier $\hat{\eta}_s$. The former difference for applications is not often that important, but the latter together with the non-absolute and non-uniform convergence make estimating this kind of sums a difficult task. One can find in the literature truncated versions of (4.4) with an error term (*cf.* (4.18) of [**62**]), the problem is that this error term blows up close to the integer numbers. This is due to the *Gibbs phenomenon*, depicted in figure 4.2 (see also §II.9 of [**98**]). Luckily for us, it is possible to perturb slightly the Fourier coefficients of ψ to obtain a finite trigonometric polynomial which approximates well ψ while staying either above or either below of this function for all x:

[2]The choice of $\mathbb{R}^2$ is made for the sake of simplicity. All the heuristics presented will still be valid in more than two dimensions.

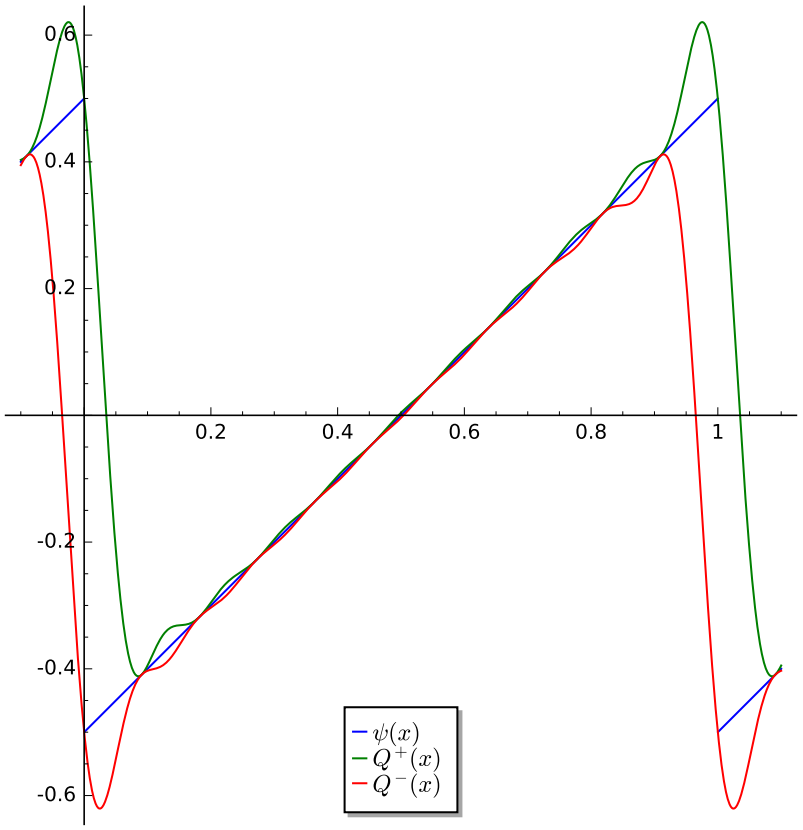


FIGURE 4.3. The saw-tooth function ψ and the Vaaler-Beurling polynomials Q^+ and Q^- of degree 10.

PROPOSITION 4.2. *For every integer $M \geq 0$ there exists trigonometric polynomials $Q^{\pm}(x) = \sum_{|m|\leq M} a_m^{\pm} e(mx)$ such that $Q^-(x) \leq \psi(x) \leq Q^+(x)$ with $a_0^{\pm} \ll M^{-1}$ and $a_m^{\pm} \ll m^{-1}$ for $0 < |m| \leq M$.*

A particular construction of such $Q^{\pm}$ was given by Vaaler and Beurling, which also have the interesting property of being extremizers. Indeed, among all trigonometric polynomials of degree at most M staying above (resp. under) ψ, Q^+ (resp. Q^-) is the one which minimizes $\left| \int_0^1 Q^+ \right|$. This result is stated as "Vaaler's lemma" in §1.2 of [**78**], and proved in §1.3 of the same book.[3] The polynomials, for $K = 10$ are shown in figure 4.3.

Now we can make the previous argument rigorous, as

$$\mathcal{N} - \sum_{|x|\leq x_0} f(n) \leq \sum_{|x|\leq x_0} Q^+(f(n)) \ll \frac{x_0}{M} + \sum_{0\neq|m|\leq M} \frac{1}{m} \left| \sum_{|n|\leq x_0} e(mf(n)) \right|,$$

and a similar formula for Q^-. Note we have lost the cancellation in m, but in our applications it will not be important, and moreover the coefficients of $Q^{\pm}$ are explicit (see §1.2 of [**78**]) should one need finer control over this.

[3] In this book the saw-tooth function ψ is modified to take the value 0 at the integer numbers, which makes the identity (4.4) hold for every $x \in \mathbb{R}$.

4.4. The van der Corput method

Suppose we want to estimate $S = \sum_{n=0}^{N} e(\phi(n))$ for some reasonably good function ϕ. The van der Corput method provides two procedures which transform the sum S into a different exponential sum, with the hope that the new exponential sum will be shorter and therefore easier to estimate. These procedures are called *process A* and *process B*, and will be informally discussed in what follows. Although there is an algorithm to decide the optimal sequence of processes A and B to apply to S under quite strong hypothesis on ϕ (see §5 of [**38**]), verifying these hypothesis is often nontrivial. When these hypothesis are not met the van der Corput method has to be combined with other methods of estimating exponential sums. In fact, some of these other methods are known to yield, for some particular sums, results beyond what is possible by purely applying A and B processes (see §7 of [**38**]). The story gets even more convoluted if the exponential sum depends on several variables, as one can either apply multidimensional methods to the whole sum or unidimensional methods to the inner-most sum. Long story short, there is no general recipe, making the estimation of exponential sums kind of an art.

The interested reader will find rigorous proofs of the two processes and many applications in Graham and Kolesnik's book [**38**] and chapter 8 of Iwaniec and Kowalski's book [**62**]. Also for simplicity we will apply the arguments directly to S, but in practice ϕ usually behaves like some power of n and therefore it is a better idea to divide the domain of the sum S diadically and estimate instead sums of the form $\sum_{n\asymp N} e(\phi(n))$.

The process B essentially consists in performing Poisson summation on the exponential sum. To do this let χ be a mollifier smooth function having the value 1 in $[0, N]$ and 0 outside $[-1/2, N + 1/2]$. Then

$$S = \sum_{n\in\mathbb{Z}} \chi(n)e(\phi(n)) \approx \sum_{n\in\mathbb{Z}} \int_0^N e(\phi(x) - nx)\, dx.$$

Now this is not an exponential sum anymore, but we can apply the stationary phase principle explained above to estimate the integral under reasonable assumptions. The phase becomes stationary when $\phi'(x) = n$, which occurs at most once if ϕ' is injective, in particular if we assume $\phi'' > 0$. Let x_n be the point satisfying $\phi'(x_n) = n$ if any. If no such x_n exists or $x_n \notin [0, N]$ then the integral is negligible because it has a lot of cancellation. Otherwise (*cf.* lemma 3.4 of [**38**]),

$$\int_0^N e(\phi(x) - nx)\, dx \approx \frac{e(\phi(x_n) - nx_n + 1/8)}{\sqrt{\phi''(x_n)}}.$$

Hence

$$S \approx \sqrt{i} \sum_{x_n\in[0,N]} \frac{e(\phi(x_n) - nx_n)}{\sqrt{\phi''(x_n)}}. \tag{4.5}$$

This can now be summed by parts to remove the smooth factor $1/\sqrt{\phi''(x_n)}$, and S may be bounded in terms of shorter sums $\sum e(\phi(x_n) - nx_n)$. The resulting sum therefore has length at most $\phi'(N)-\phi'(0)+1$. This process is therefore advantageous when the variation of ϕ' is small. This is the heuristic, at least, because if the variation were too small, that would mean ϕ would be almost linear and the sum of the geometric series shows the above result is too good to be true. Indeed the size

of

$$\sum_{n=0}^{N} e(An+B) = e(B)\frac{e\big(A(N+1)\big)-1}{e(A)-1} \tag{4.6}$$

is close to N when A is close to being an integer. Of course, some error terms we have neglected blow up when $\phi'' \approx 0$, and so does every term of the resulting sum, and hence we would better say that process B is advantageous when ϕ'' is "reasonably small". A rigorous statement of process B is contained in lemma 3.6 of [**38**] (see also exercise 3 of §8.3 of [**62**] for a version with summands of arbitrary modulus).

When ϕ'' is too big we can usually reduce its size by process A, also called performing a *Weyl step*. The idea is simple: if we square $|S|$,

$$|S|^2 = \sum_{0\le n,m\le N} e\big(\phi(n)-\phi(m)\big)$$

and $\phi(n)-\phi(m) = (n-m)\phi'(x_{n,m})$. If N is small, so must be $n-m$, and for many reasonable functions the derivative has less variation than the function itself. Hence we are essentially replacing ϕ'' by ϕ'''.

Usually N is not so small, and the trick is to break the sum S into smaller sums and square each of them. For simplicity suppose $H \mid N$ and the sum S runs from $n=0$ to $n=N-1$. Then by Cauchy-Schwarz,

$$|S|^2 \le \frac{N}{H}\sum_{k=0}^{N/H-1}\left|\sum_{n=kH}^{(k+1)H-1} e\big(\phi(n)\big)\right|^2 = \frac{N}{H}\sum_{k=0}^{N/H-1}\sum_{kH\le n,m\le (k+1)H-1} e\big(\phi(n)-\phi(m)\big).$$

Writing $n = m+\ell$ and separating by cases on whether ℓ is positive, negative or zero,

$$|S|^2 \le \frac{N^2}{H} + \frac{N}{H}\sum_{1\le\ell\le H}\left|\sum_{k=0}^{N/H-1}\sum_{kH\le m,m+\ell\le (k+1)H-1} e\big(\phi(m+\ell)-\phi(m)\big)\right|.$$

Now $\phi(m+\ell)-\phi(m) \approx \ell\phi'(m)$ for H reasonably small. This is done with a prettier approach in §2.3 of [**38**]. Note that the length of the sum can be though to stay invariant (the sum over ℓ is an average as we are dividing over H; the other ones have combined length N). Nevertheless, even if we were able to prove square-root cancellation in the resulting sum for $H = N$ this would result in $|S| \ll N^{3/4}$: performing a Weyl step has the cost of, at best, halving any power-savings we can get from the resulting exponential sum.

If on the other hand we find that process B fails because ϕ'' is too small, then we can still prove that the sum has cancellation as long as ϕ' stays away from the integers (note ϕ' plays the same role as A in (4.6)). This is usually called the Kuzmin-Landau lemma, which we prove next.

Proposition 4.3 (Kuzmin-Landau). *If ϕ continuously differentiable, ϕ' is monotone and $\|\phi'\|_{\mathbb{Z}} \ge \lambda > 0$, then $S \ll \lambda^{-1}$.*

Proof. (Adapted from theorem 2.1 of [**38**], argument originally due to Mordell [**79**]) By conjugating S if necessary we may assume that ϕ' is increasing, and by substituting $\phi(n)$ by $\phi(n)-kn$ for an appropriate integer k that $\lambda \le \phi' \le 1-\lambda$.

If S was truly a geometric series, $\phi(n) = An$, then writing $e(An) = \big(e(A(n+1)) - e(An)\big)/\big(e(A)-1\big)$ would telescope the series. We follow the same idea and

write

$$S = \sum_{n=0}^{N-1} \Big(e\big(\phi(n+1)\big) - e\big(\phi(n)\big)\Big) C_n + e\big(\phi(N)\big) \quad \text{where} \quad C_n = \frac{1}{e\big(\phi(n+1) - \phi(n)\big) - 1}.$$

Summing by parts,

$$\begin{aligned} S &= \sum_{n=1}^{N-1} e\big(\phi(n)\big)\big(C_{n-1} - C_n\big) + e\big(\phi(N)\big)\big(C_{N-1} + 1\big) - e\big(\phi(0)\big) C_0 \\ &\leq \sum_{n=1}^{N-1} \big|C_{n-1} - C_n\big| + \big|C_0\big| + \big|C_{N-1}\big| + 1. \end{aligned}$$

Note we have $1/(e(\eta) - 1) = -\frac{1}{2}(1 + i \operatorname{cotan}(\pi\eta))$, and hence writing $\eta_n = \phi(n+1) - \phi(n)$,

$$S \leq \frac{1}{2} \sum_{n=1}^{N-1} \Big| \frac{1}{\tan(\pi\eta_{n-1})} - \frac{1}{\tan(\pi\eta_n)} \Big| + \frac{1}{\big|\tan(\pi\eta_0)\big|} + \frac{1}{\big|\tan(\pi\eta_{N-1})\big|} + 2.$$

By the mean value theorem, η_n is an increasing function of n, lying between λ and $1 - \lambda$. Hence the series telescopes and the bound $|\operatorname{cotan}(\pi\eta_n)| \ll \lambda^{-1}$, valid for all n, shows $S \ll \lambda^{-1}$. □

The simplest van der Corput estimate is obtained by applying process B to S and then estimating the resulting sum term by term. Assume $\phi''(x) \asymp \lambda$. If (4.5) were true as is, we would obtain $S \ll (N\lambda + 1)\lambda^{-1/2}$. This bound is still true, even if we rigorously take into account the neglected error terms (*cf.* lemma 3.6 of [**38**]).

Proposition 4.4 (van der Corput's lemma). *If ϕ has two continuous derivatives and $0 < \lambda \leq |\phi''(x)| \leq \alpha\lambda$ then $S \ll \alpha N \lambda^{1/2} + \lambda^{-1/2}$.*

There is a much simpler proof of this result which we can provide here, as it does not require the previous detour through process B.

Proof. (Adapted from theorem 2.2 of [**38**]) The idea is to apply the Kuzmin-Landau bound whenever possible. Conjugating the series, if necessary, we may assume ϕ'' is everywhere positive, and hence ϕ' is monotone increasing. Note also we may assume $\lambda \leq 1$, as otherwise the trivial estimation provides a better bound. Fix $\delta > 0$ to be chosen later, and let $\Omega = \{0 \leq x \leq N : \|\phi'\|_{\mathbb{Z}} \geq \delta\}$. By the mean value theorem, $\phi'(N) - \phi'(0) \leq N\alpha\lambda$, and therefore Ω consists of at most $N\alpha\lambda + 2$ intervals. Hence

$$\sum_{n \in \Omega} e\big(\phi(n)\big) \ll (N\alpha\lambda + 2)\delta^{-1}.$$

On the other hand, the complement of Ω in $[0, N]$ contains at most $N\alpha\lambda + 3$ intervals, delimited by the points where $\phi'(x) = n \pm \delta$ or the limits of the interval $[0, N]$. By the mean value theorem, each of these have length $(\phi')^{-1}(n + \delta) - (\phi')^{-1}(n - \delta) = 2\delta/\phi''(\xi) \leq 2\delta\lambda^{-1}$. Hence the trivial estimation yields

$$\sum_{n \notin \Omega} e\big(\phi(n)\big) \leq (N\alpha\lambda + 3)(2\delta\lambda^{-1} + 1).$$

Choosing $\delta = \lambda^{1/2}$ we obtain the right bound. □

The general strategy of the van der Corput method can therefore be summarized as applying process A until the second derivative is in the range of either applying Kuzmin-Landau or process B, and if in the latter case repeat. Note also that Kuzmin-Landau extracts the cancellation from ϕ', the van der Corput lemma above from ϕ'', this very same lemma after a process A would extract the cancellation from ϕ''', etc. The same method can also be understood as expanding ϕ by its Taylor series in short intervals, and extracting the cancellation from the first monomial with a coefficient of the right size (see §8.2 of [**62**]). In fact, the Weyl step was first used by Weyl in the case where ϕ is a polynomial, to reduce S to a geometric series and show that if the leading coefficient is not a rational with denominator dividing $(\deg \phi)!$ then $S \ll N^{1-\gamma}$ for some $\gamma > 0$. This together with Weyl's criterion (see §2 of the introduction) shows that the sequence $\{\phi(n)\}_{n\in\mathbb{Z}}$ is equidistributed modulo 1 if the leading coefficient of ϕ is irrational.

In chapter 6 we will require another use of the Weyl step. Imagine we have an exponential sum in two variables $S = \sum_{n,m} w(n)e(\phi(n,m))$ where w is a function which changes size wildly, for example an arithmetic function, but is bounded above by W. We can still consider the exponential sum in m and apply the method above, but here the cancellation obtained by the van der Corput method might be very poor. For example if the sum in n is much longer than the one in m, any power-savings we can get are probably going to have a bigger effect if we can get them in the n variable. Although we cannot use the cancellation in n directly because we do not know w well enough, we can use the variation of ϕ with respect to n to show that the cancellation in m must be most of the time smaller than we would expect. Squaring $|S|$ and using Cauchy-Schwarz,

$$\begin{aligned}
|S|^2 &\leq NW^2 \sum_{n\leq N} \left| \sum_{m\leq M} e(\phi(n,m)) \right|^2 \\
&= NW^2 \sum_{n\leq N} \sum_{m\leq M} \sum_{|\ell|\leq M} e(\phi(n,m+\ell) - \phi(n,m)) \\
&= N^2MW^2 + NW^2 \Re \sum_{n\leq N} \sum_{m\leq M} \sum_{0<\ell\leq M} e(\phi(n,m+\ell) - \phi(n,m)) \\
&\ll N^2MW^2 + NW^2 \sum_{m\leq M} \sum_{0<\ell\leq M} \left| \sum_{n\leq N} e(\phi(n,m+\ell) - \phi(n,m)) \right|.
\end{aligned}$$

As a toy example, if we assume we have the same power-savings in the original sum in m and in the inner sum of the Weyl step, admitting bounds respectively of $M^{1-\gamma}$ and $N^{1-\gamma}$, $\gamma \leq 1/2$, then for $N \geq M^2$ the Weyl step produces the best bound.

CHAPTER 5

Lattice points in elliptic paraboloids

This chapter focuses in the results contained in the article "Lattice points in elliptic paraboloids" [**20**], joint work with F. Chamizo.

5.1. Main results

The classical and most paradigmatic lattice point counting problems —Gauss' circle problem and Dirichlet's divisor problem— correspond to two of to the simplest conics, the circle $x^2+y^2\leq 1$ and the hyperbola $xy\leq 1$. More arbitrary rational ellipses and hyperbolas appear when deriving Dirichlet's class number formula (see §3 of the introduction and chapter 6 of [**23**]). For all these problems (and probably for much more arbitrary bidimensional shapes) the best known result is Bourgain and Watt's $\alpha_K\leq 517/824$ (*cf.* §4.1).

The remaining conic, the parabola, did not attract much attention until very recently. As with the hyperbola, it has the problem of not being closed, and therefore has to be somehow truncated. Popov in 1975 was the first to consider a parabolic region, counting the number of points with integer coordinates in the region $0\leq x\leq A$, $0\leq y\leq x^2/B$ where B is an integer, A is an arbitrary positive real number, $B\leq A$ and the points lying in the x-axis are counted with weight $1/2$.[1] For this problem he obtained an error term of size $O(A^{1/2})$, not depending on B. Note that by setting $A=RA_0$ and $B=RB_0$ we obtain $\alpha_K\leq 1/2$ under the formalism of §4.1: for this region it is surprisingly simple to obtain the conjecture that is currently out of reach for the circle and the hyperbola. In fact, we will provide a version of Popov's proof, further simplified by the use of the Vaaler and Beurling polynomials and standard bounds for quadratic exponential sums, in section 5.2 below. For simplicity we will phrase the result in terms of the number of points with integer coordinates in the region

$$\mathcal{P}_2=\{|y|\leq c-(x-\beta)^2\}, \tag{5.1}$$

but the same proof works for the regions considered by Popov.

An elementary argument shows that $\alpha_{\mathcal{P}_2}\geq 1/2$ when $\beta=0$ and c is a rational number, an hence for these cases $\alpha_{\mathcal{P}_2}=1/2$. We will revisit this result also in §5.2, and find for the particular choice $\beta=0$ and $c=1$ an exact formula for the error term $\mathcal{E}(R)$ in terms of a sum of L-functions evaluated at 1. In particular, this will show for this particular choice of $\mathcal{P}_2$ that

$$\mathcal{E}(R)=\Omega_-\big(R^{1/2}\exp(a\sqrt{\log R}/\log\log R)\big)\qquad\text{for any}\quad a<\sqrt{2}. \tag{5.2}$$

[1] This is very common when counting lattice points in regions where there is a fixed straight edge. Usually the region formed by adjoining the reflection through the straight edge has better properties in terms of curved boundary, and therefore this one has a better chance of having a small error exponent α_K. Since the points in the straight edge are shared between the two halves, if one wants to obtain a small error exponent for each of the parts these points must be necessarily counted with weight $1/2$. The same phenomenon underlies the coefficient $1/2$ appearing in the Euler-Maclaurin formula (*cf.* §**A**.4).

The same proof also generalizes to $c \in \mathbb{Z}$. This result also contrasts with the literature for the circle and the hyperbola, where it is not known whether $\mathcal{E}(R) = \Omega(R^{1/2}(\log R)^{1/2+\epsilon})$ for any $\epsilon > 0$, but it is thought to be unlikely [**40, 90**].

In higher dimensions the balls and ellipsoids have been throughout studied. In particular, the conjecture is known in the rational case for $d \geq 4$ and in the irrational for $d \geq 5$. In three dimensions the best known result for the ball is Heath-Brown's $\alpha_K \leq 21/16$ also valid for rational ellipsoids (*cf.* §4.1). The same error exponent also holds for the average of the class number, which can be regarded as a lattice point counting problem in a three-dimensional region delimited by hyperboloids [**18**].

Again, one can find very little literature regarding parabolic regions. The natural analogue of the set $\mathcal{P}_2$ defined above is the elliptic paraboloid

$$\mathcal{P} = \{(\vec{x}, y) \in \mathbb{R}^{d-1} \times \mathbb{R} \,:\, |y| \leq c - Q(\vec{x} + \vec{\beta})\}, \tag{5.3}$$

where Q is a positive definite quadratic form, $\vec{\beta}$ is a fixed vector in $\mathbb{R}^{d-1}$ and c a positive constant. The particular case $\vec{\beta} = 0$ was considered in a slightly different form by Krätzel in [**71, 72**], where he showed that Hlawka's result $\alpha_{\mathcal{P}} \leq d(d-1)/(d+1)$ holds in general and moreover that the conjecture $\alpha_{\mathcal{P}} \leq d-2$ holds under the strong assumptions $d \geq 5$ and Q either rational or diagonal. Partial results were given under weaker rationality assumptions in terms of the coefficient matrix $A = (a_{ij})$ of Q. In particular, Krätzel obtained $\alpha_{\mathcal{P}} \leq d - 5/3$ for $d \geq 3$ as long as a_{12}/a_{11}, $a_{22}/a_{11} \in \mathbb{Q}$. We improve these results:

THEOREM 5.1. *If a_{12}/a_{11}, $a_{22}/a_{11} \in \mathbb{Q}$ then the inequality $\alpha_{\mathcal{P}} \leq d-2$ holds for any $d \geq 3$.*

As with the parabola we will also provide Ω-results for the error term in the case $\vec{\beta} = 0$, $c \in \mathbb{Q}$ and Q rational, which show that theorem 5.1 is sharp under these hypotheses.

THEOREM 5.2. *Suppose $\vec{\beta} = 0$, $c \in \mathbb{Q}$ and Q rational. Then for $d \geq 3$ we have $\mathcal{E}(R) = \Omega(R^{d-2}\eta(R))$ where*

$$\eta(R) = \begin{cases} \exp\left(a\frac{\log R}{\log\log R}\right) & \text{for any } a < \log 2 \text{ when } d = 3, \\ \log\log R & \text{when } d = 4, \\ \sqrt{\log\log R} & \text{when } d = 5, \\ 1 & \text{when } d \geq 6. \end{cases}$$

In particular, $\alpha_{\mathcal{P}} = d - 2$.

Note that in theorem 5.1 no assumptions are imposed on the remaining coefficients, and therefore this result extends the upper bound $\alpha_{\mathcal{P}} \leq d-2$ not only to $d = 3$, 4 and $\vec{\beta} \neq 0$, but also to a wider family of higher-dimensional paraboloids for which Krätzel's result does not apply. The key step in the proof are the bounds we obtained in §2.7 employing what can be considered a toy version of the circle method, as these can be applied to the associated exponential sum because for the region $\mathcal{P}$ it essentially corresponds to a truncated theta series. Bounds this precise are out of reach for the exponential sums arising in most lattice point problems, and this accounts for the striking difference between our theorem and what is currently known for ellipsoids and hyperboloids. In fact, to the best of my knowledge, theorem 5.1 constitutes the first sharp result for a lattice point problem in three dimensions.

Note also that (5.2) and theorem 5.2 show that when $2 \le d \le 5$ the ϵ in the bound for the error term $\mathcal{E}(R) = O(R^{\alpha_{\mathcal{P}}+\epsilon})$ cannot be dropped (for $d = 2$ under the stronger assumption $c \in \mathbb{Z}$). In contrast, when $d \ge 6$ the lattice point discrepancy is actually $O(R^{d-2})$, as shown by applying Euler-Maclaurin summation to the corresponding asymptotics for the number of lattice points in the dilated $(d-1)$-dimensional ellipsoid $\{Q(\vec{x}) \le 1\}$ (see §4.1). For irrational paraboloids our method does not provide an answer as to whether the ϵ is really necessary.

5.2. The parabola

Let us start with the short proof of Popov's result for the region (5.1). Since the number of points with integer coordinates in $R\mathcal{P}_2$ does not vary if we displace this set an integer amount in the x direction, we may assume that $R\beta \in [0,1)$. Let $f(x) = c - (x-\beta)^2$, and note that we have

$$\begin{aligned} \frac{1}{2}\mathcal{N}(R) &= \sum_{f(n/R)\ge 0} \left(\lfloor Rf(n/R) \rfloor + \frac{1}{2} \right) \\ &= R \sum_{f(n/R)\ge 0} f(n/R) - \sum_{f(n/R)\ge 0} \psi(Rf(n/R)). \end{aligned} \tag{5.4}$$

The first sum is easily seen to equal $R|\mathcal{P}_2| + O(1/R)$ by an application of the Euler-Maclaurin formula. Using the Vaaler–Beurling polynomials of degree $\lfloor R^{1/2} \rfloor$ (proposition 4.2), this implies

$$\mathcal{E}(R) \ll R^{1/2} + \sum_{0<|m|\le R^{1/2}} \frac{1}{m} \left| \sum_{f(n/R)\ge 0} e\left(\frac{m}{R} n^2 - 2\beta mn \right) \right|.$$

The exponential sum runs over the integers in the interval $[R\beta - Rc^{1/2}, R\beta + Rc^{1/2}]$, which may be replaced with $[-Rc^{1/2}, Rc^{1/2}]$ at the cost of adding and subtracting a finite number of terms. By the Hardy-Littlewood bound (2.11), which is also valid with a linear term inside the exponential, and which admits a very simple proof if we add the extra error term $R^{1/2}\log R$ (see §8.2 of [**62**]),

$$S_m = \sum_{|n|\le Rc^{1/2}} e\left(\frac{m}{R} n^2 - 2\beta mn \right) \ll \frac{R}{q_m^{1/2}} + R^{1/2}\log R$$

where p_m/q_m is a rational satisfying

$$\left| 2\frac{m}{R} - \frac{p_m}{q_m} \right| \le \frac{1}{q_m \lfloor Rc^{1/2} \rfloor} \quad \text{with} \quad q_m \le \lfloor Rc^{1/2} \rfloor, \tag{5.5}$$

which is guaranteed to exist by Dirichlet's approximation theorem.[2] Assume first that $c \ge 1$, and note that this together with the condition (5.5) ensures $p_m \ne 0$ and $q_m \asymp Rp_m/(2m)$. Hence

$$\sum_{0<|m|\le R^{1/2}} \frac{|S_m|}{m} \ll R^{1/2} \sum_{0<|m|\le R^{1/2}} \frac{1}{(mp_m)^{1/2}} + R^{1/2}(\log R)^2. \tag{5.6}$$

[2]When applying Dirichlet's approximation theorem to a random real, $|x - p/q| \le (qN)^{-1}$ with $q \le N$, typically $q \gg N^{1-\epsilon}$. This follows from the fact that $\sum_{q\le N} q^{-1}\phi(q) \asymp N$ where ϕ stands for Euler's totient function (theorem 330 of [**46**]), by showing that the area covered by the intervals $[x - 1/(qN), x + 1/(qN)]$ for $q \le N\delta$ tends linearly to zero as $\delta \to 0^+$. Note that if we always had $q_m \gg R^{1-\epsilon}$ in the proof above the result would be immediate. In some sense, the remaining part of the proof consists in showing that this is true for $x = m/R$ in the sense of the given average.

Now, $1 \leq p_m \ll m$ by $p_m \asymp 2mq_m/R$ and $\#\{m : p_m = p\} \leq R^\epsilon$ for each p, as $p_m = p$ implies $|2mq_m - Rp| \leq 2$ by (5.5) and therefore m divides some integer in the interval $[Rp-2, Rp+2]$. Hence

$$\sum_{1\leq m\leq M} \frac{1}{p_m^{1/2}} \leq R^\epsilon \sum_{1\leq p\ll M} \frac{1}{p^{1/2}} \ll R^\epsilon M^{1/2}$$

and summing by parts in (5.6) we obtain $\mathcal{E}(R) \ll R^{1/2+\epsilon}$.

The case $c < 1$ remains. The only thing that breaks down in this case is that when $2m/R$ is too small ($p_m = 0$) the Hardy-Littlewood bound as presented only provides the trivial estimation. This is because the Farey dissection is too rough and does not distinguish $2m/R$ from zero. The solution is simple. A trivial modification of the proof provided in [**62**] shows that the same bound holds if we replace (5.5) by

$$\left|2\frac{m}{R} - \frac{p_m}{q_m}\right| \leq \frac{1}{q_m\lfloor KRc^{1/2}\rfloor} \quad \text{with} \quad q_m \leq \lfloor KRc^{1/2}\rfloor$$

for some fixed $K > 0$. It suffices to take $K = c^{-1/2}$.

One may find surprising that it is possible to obtain the conjecture only applying Poisson summation in one variable. One heuristic explanation could be the following: the exponential sum —up to a linear term— corresponds to a truncated version of Jacobi's theta function evaluated at m/R. Applying Poisson summation in the n variable then it is essentially equivalent to the transformation formula $\theta(z) = (-iz)^{-1/2}\theta(-1/z)$ (*cf.* §2.2). But either before or after the Poisson summation the sum left to estimate is a truncated version of a modular form, and these we know quite well. Since the Poisson summation does not increase the overall cancellation of the sum, it just makes it easier to spot, it seems plausible that in this case it is unnecessary.

The same proof will be essentially repeated in the next section for the case of a paraboloid in $\mathbb{R}^3$, but this time using the bounds of proposition 2.13 instead of Hardy-Littlewood's bound, as the exponential sum will be a truncated version of θ^2 or a similar theta function of weight 1. The same heuristics are valid in this case, and will allow us to use again the shortcut of the Vaaler and Beurling polynomials.

Note that in the proof of proposition 2.13, which was based on a toy version of the circle method, the main contribution comes from the piece of the integral lying over the Ford circle associated to the rational p/q, close to x. The expansion on this cusp (theorem 2.4) is obtained by transforming the modular form via the slash operator by a matrix sending p/q to ∞. This matrix is essentially applying S and T to undo the continued fraction expansion of p/q until we obtain ∞ (cf. §1.3). Since applying S essentially amounts to applying Poisson summation, morally it should not matter if instead we directly transform the original truncated exponential sum via translations $x \mapsto x+1$ and the process B of the van der Corput method (§4.4), in a way dictated by the continued fraction expansion of p/q. This is in fact the idea behind Hardy and Littlewood's work [**45**], where they do this directly with the truncated series of the θ function, as for weight smaller than one the circle method has problems of convergence. This also means that one can bypass modular forms altogether and use the van der Corput method directly to obtain the same result, but of course the machinery developed in chapters 1 and 2 is a very convenient way of carrying this out with little effort.

Aside from proving the upper bound for $\alpha_{\mathcal{P}_2}$, Popov in his article [**81**] also remarked that when $c \in \mathbb{Q}$ and $\beta = 0$ one had $\alpha_{\mathcal{P}_2} = 1/2$. For this it suffices to show

that we can find arbitrarily large values of R for which there are at least $R^{1/2}$ points on the boundary of $R\mathcal{P}_2$, as then $\mathcal{E}(R)$ will necessarily have jump discontinuities of this size and we will have $\mathcal{E}(R) = \Omega(R^{1/2})$. If $c = a/b$ we may take $R = b^2N^2$ for any large integer N as then all the points in the boundary of $R\mathcal{P}_2$ whose abscissa is an integer multiple of bN have integer coordinates. There are approximately $2(a/b)^{1/2}R^{1/2}$ such points.

In fact, it is possible to give an exact formula for $\mathcal{E}(R)$ when both c and the dilation factor R are integers, from which we can prove the stronger Ω-result (5.2). This is done by substituting the saw-tooth function ψ by its Fourier series directly instead of using the Vaaler and Beurling polynomials, and then evaluating explicitly the resulting quadratic Gauss sums. In this way we obtain a formula relating the error term for this lattice point problem to the class number associated to a family of imaginary quadratic fields. This relation with the class number was pointed out by Professor Antonio Córdoba in the early 90's while he was the Ph.D. advisor of F. Chamizo.

When c is an arbitrary rational number one might be able to obtain similar Ω-results by employing estimates of incomplete quadratic Gauss sums (see [**74**]).

For the sake of simplicity we are only going to consider the case $\beta = 0$ and $c = 1$. Hence for the rest of this section we will assume that $\mathcal{P}_2$ is determined by the inequality $|y| \leq 1 - x^2$.

THEOREM 5.3. *Let N be an odd positive integer and let N^* be the largest square dividing N. Then*

$$\mathcal{N}(N) = |\mathcal{P}_2|N^2 + \frac{1}{3} + 2\sqrt{N^*} - \frac{4}{\pi} \sum_{\substack{d|N \\ d\equiv 3 \ (4)}} \sqrt{d}L(1, \chi_{-d})$$

where $L(1, \chi_{-d})$ is the L-function corresponding to the Kronecker symbol $\chi_{-d} = \left(\frac{-d}{\cdot}\right)$.

With some effort the result can be extended, with modifications, to cover the even case.

Two particular cases of theorem 5.3 deserve special attention, and will be used to obtain the aforementioned one-sided Ω-results.

COROLLARY 5.4. *If the prime factors of N are of the form $4k + 1$, then*

$$\mathcal{E}(N) = \frac{1}{3} + 2\sqrt{N^*}.$$

COROLLARY 5.5. *If N is squarefree then*

$$\mathcal{E}(N) = \frac{7}{3} - 4 \sum_{\substack{d|N \\ d\equiv 3 \ (4)}} \omega_d h(-d)$$

where $h(-d)$ is the class number of the integer ring of $\mathbb{Q}(\sqrt{-d})$ and $\omega_d = 1$ except for $\omega_3 = 1/3$.

PROOF. Apply Dirichlet class number formula (**I**.17) in theorem 5.3 for the fundamental discriminant $-d$. □

PROOF OF THEOREM 5.3. By (5.4),

$$\mathcal{N}(N) = 2\sum_{n=-N}^{N}\Big(N - \frac{n^2}{N}\Big) - 2\sum_{n=-N}^{N}\psi\Big(-\frac{n^2}{N}\Big).$$

The first sum is $(4N^2-1)/3$ and the area is $|\mathcal{P}_2| = 8/3$. Then

$$\mathcal{E}(N) = -\frac{2}{3} - 2\sum_{n=-N}^{N}\psi\Big(-\frac{n^2}{N}\Big) = \frac{1}{3} - 4\sum_{n=1}^{N}\psi\Big(-\frac{n^2}{N}\Big).$$

The Fourier series of ψ (4.4) converges to $\psi(x)$ when x is not an integer and to 0 otherwise. Hence

$$\psi(x) = \Im\sum_{m=1}^{\infty}\frac{e(-mx)}{\pi m} + \begin{cases}0 & \text{if } x\notin\mathbb{Z},\\ -1/2 & \text{if } x\in\mathbb{Z}.\end{cases}$$

Note that N divides n^2 exactly $\sqrt{N^*}$ times in the range $1\le n\le N$, and hence

$$\mathcal{E}(N) = \frac{1}{3} + 2\sqrt{N^*} - \frac{4}{\pi}\sum_{m=1}^{\infty}\frac{1}{m}\Im G(m;N) \tag{5.7}$$

where $G(m;N)$ is the quadratic Gauss sum $\sum_{n=1}^{N} e(mn^2/N)$. Let $d_m = N/\gcd(m,N)$, the evaluation of $\Im G(m;N)$ reads (see exercise 4 of §3.5 of [**62**])

$$\Im G(m;N) = \begin{cases}0 & \text{if } d_m \equiv 1 \pmod 4,\\ \frac{N}{\sqrt{d_m}}\Big(\frac{md_m/N}{d_m}\Big) & \text{if } d_m \equiv 3 \pmod 4.\end{cases}$$

When d_m is fixed and $1\le m\le M$, the quantity md_m/N runs over all positive integers coprime to d_m in the range $1,\dots,\lfloor Md/N\rfloor$. Hence substituting in (5.7) we have

$$\mathcal{N}_2(N) - |\mathcal{P}_2|N^2 = \frac{1}{3} + 2\sqrt{N^*} - \frac{4}{\pi}\sum_{\substack{d\mid N\\ d\equiv 3 \ (4)}}\sqrt{d}\sum_{m=1}^{\infty}\frac{1}{m}\Big(\frac{m}{d}\Big).$$

By the quadratic reciprocity law for the Jacobi-Kronecker symbol (exercise 3 and (3.43) of §3.5 of [**62**]), the innermost sum equals $L(1,\chi_{-d})$. □

From corollaries 5.4 and 5.5 the following refinement of the Ω-result $R^{1/2}$ is immediate.

PROPOSITION 5.6. *The error term satisfies*

$$\mathcal{E}(R) = \Omega_+\big(R^{1/2}\big) \quad \textit{and} \quad \mathcal{E}(R) = \Omega_-\big(R^{1/2}\log\log R\big).$$

PROOF. The first statement follows by taking N a square in corollary 5.4. For the second one we remark that the main result of [**4**] asserts that there are infinitely many primes $p\equiv 3 \pmod 4$ satisfying $h(-p)/\sqrt{p} \gg \log\log p$. It suffices to take $N = p$ for any such prime p in corollary 5.5. □

The upper bound $h(-d)/\sqrt{d} \ll \log\log d$ is known to hold under the generalized Riemann hypothesis [**75**]. Any hope to obtain a better Ω_--result from corollary 5.5 therefore must take advantage of the sum of class numbers, and for this we need uniform lower bounds over certain families of discriminants. Fortunately Heath-Brown proved an astonishing result that, in some way, shows the absence of exceptional zeros for large multiples of some primes in a fixed set [**48**]. Even more astonishing

is the short and elementary proof of this fact. In its original form the result claims that if S is a fixed set of more than 505^2 odd primes then for any sufficiently large integer d there exists a prime $p_d \in S$ satisfying $L(1, \chi_{-p_d d}) \gg (\log d)^{-1/9}$. Since the original text seems to be hard to find, we provide here a version with a slightly more general statement. This version can also be found stated without proof by Blomer in [**8**].

PROPOSITION 5.7 (HEATH-BROWN). *Fix $\epsilon > 0$ and let S be a set of primes congruent to 3 modulo 4, of cardinality $\#S > (1 + 2/\epsilon)^4$. There is an integer $N > 0$ such that for every $n \geq N$, $n \equiv 1 \pmod 4$, there is some $p_n \in S$ satisfying*

$$L(1, \chi_{-np_n}) \gg (\log n)^{-\epsilon}.$$

The hypotheses regarding the congruence classes of n and the primes in S modulo 4 are only included for the sake of simplicity, to ensure that n, $-p$ and $-np$ are fundamental discriminants and therefore the Kronecker symbol χ_d is well-defined for them.

PROOF. We are going to assume $L(1, \chi_{-np}) \leq \big(\log np\big)^{-\epsilon}$ for every $p \in S$ and from here deduce that the set S must be smaller than $(1+2/\epsilon)^4$. All the implicit constants in the argument may depend on S.

It is convenient to translate the bound on $L(1, \chi_{-np})$ into a bound for $L(\sigma, \chi_{-np})$ for some $\sigma > 1$, as here the Euler product converges well. For this we use the mean value theorem. Note that $L'(\sigma, \chi_{-np}) \ll (\log n)^2$ (see (11) of chapter 14 of [**23**]) and hence

$$L(\sigma_0, \chi_{-np}) = L(1, \chi_{-np}) + O\big(|\sigma_0 - 1|(\log n)^2\big) \ll (\log n)^{-\epsilon} \tag{5.8}$$

for $\sigma_0 = 1 + (\log n)^{-2-\epsilon}$.

Considering the Euler product of $L(\sigma_0, \chi_{-np})$ omitting the factors corresponding to the primes in S,

$$\log L(\sigma_0, \chi_{-np}) = \sum_{m \geq 1}{}' \frac{\Lambda(m)}{m^{\sigma_0} \log m} \chi_{-np}(m) + O(1) \tag{5.9}$$

where Λ is the usual von-Mangoldt function and the prime indicates we are summing only over those m coprime to $P = \prod_{p \in S} p$. Since (5.8) implies

$$\big(\log L(\sigma_0, \chi_{-np}) + O(1)\big)^2 \geq \epsilon^2\big(1 + o(1)\big)(\log\log n)^2,$$

substituting (5.9) and summing over $p \in S$,

$$\sum_{m_1 \geq 1}{}' \sum_{m_2 \geq 1}{}' \frac{\Lambda(m_1)\Lambda(m_2)}{(m_1 m_2)^{\sigma_0} \log m_1 \log m_2} \sum_{p \in S} \chi_{-np}(m_1 m_2) \geq \epsilon^2\big(1 + o(1)\big)(\log\log n)^2 \# S.$$

Using $\chi_{-np}(m) = \chi_n(m)\chi_{-p}(m)$ and dividing into classes modulo P, the left hand side is bounded above by

$$\sum_{a_1=1}^{P}{}' \sum_{a_2=1}^{P}{}' \Big| \sum_{p \in S} \chi_{-p}(a_1 a_2) \Big| L(a_1) L(a_2) \qquad \text{with} \quad L(a) = \sum_{m \equiv a \pmod P} \frac{\Lambda(m)}{m^{\sigma_0} \log m}.$$

If we drop the condition $m \equiv a \pmod P$ in the sum defining $L(a)$ then this quantity becomes $\log \zeta(\sigma_0) = (2 + \epsilon) \log\log n + o(1)$. In general the congruence condition can be detected by summing over all characters modulo P:

$$\phi(P) L(a) = \log \zeta(\sigma_0) + \sum_{\chi \pmod P} \overline{\chi}(a) \log L(\sigma_0, \chi),$$

$$= (2+\epsilon)\log\log n + O(1),$$

where ϕ stands for Euler's totient function. Hence, substituting above,

$$\frac{1}{\phi^2(P)}\sideset{}{'}\sum_{a_1=1}^{P}\sideset{}{'}\sum_{a_2=1}^{P}\Big|\sum_{p\in S}\chi_{-p}(a_1a_2)\Big| \geq \frac{\epsilon^2}{(2+\epsilon)^2}(1+o(1))\#S.$$

For any k coprime to P the equation $xy \equiv k \pmod P$ has exactly $\phi(P)$ solutions, and therefore the left hand side squared is

$$\frac{1}{\phi^2(P)}\left(\sideset{}{'}\sum_{k=1}^{P}\Big|\sum_{p\in S}\chi_{-p}(k)\Big|\right)^2 \leq \frac{1}{\phi(P)}\sum_{p_1\in S}\sum_{p_2\in S}\sideset{}{'}\sum_{k=1}^{P}\chi_{p_1p_2}(k) = \#S.$$

This, together with the previous inequality, shows $\sqrt{\#S} \leq (1+2/\epsilon)^2(1+o(1))$, a contradiction. □

Using Heath-Brown's result we can prove (5.2):

Proof of (5.2). Let S be the set of the first 3^4+1 primes $p \equiv 3 \pmod 4$ and fix an integer d_0 large enough so that the aforementioned result of Heath-Brown holds for any $d \geq d_0$. Choose $N = N'\prod_{p\in S} p$ in corollary 5.5, where N' is the product of the primes $p \equiv 1 \pmod 4$ in the interval $[d_0, x]$ for any large x. Then by the class number formula,

$$\sum_{\substack{d\mid N\\ d\equiv 3 \ (4)}} \omega_d h(-d) \gg \sum_{\substack{d\mid N'\\ d\neq 1}} \frac{\sqrt{p_d d}}{\log d} \gg \frac{\sqrt{N'}}{\log N}\prod_{p\mid N'}(1+p^{-1/2}) + o(1).$$

The result now follows by noting that by the prime number theorem in arithmetic progressions, the logarithm of the product over the primes is asymptotically $\sqrt{2\log N'}/\log\log N'$ and $N' \gg N$. □

5.3. Elliptic paraboloids

The proof of theorem 5.1 for $d = 3$ will follow closely the proof we have given of Popov's result for the parabola. In this case we can exploit the full rationality of the quadratic form Q. When $d \geq 4$ we will just slice the paraboloid into three-dimensional paraboloids and then glue the results together via the Euler-Maclaurin formula. To carry this out succesfully we need the error term to be uniformly bounded in terms of the parameters c and $\vec{\beta}$. For convenience we state here this slightly more precise version of theorem 5.1.

Theorem 5.8. *Let $\mathcal{P}$ be as in (5.3) with $d \geq 3$. Assume that the coefficient matrix $A = (a_{ij})$ of Q satisfies $a_{12}/a_{11}, a_{22}/a_{11} \in \mathbb{Q}$. Then for each fixed $\epsilon > 0$, $C > 0$,*

$$\mathcal{N}(R) = |\mathcal{P}|R^d + O\big(R^{d-2+\epsilon}\big)$$

holds uniformly for $R \geq 1$, $0 < c < C$ and $\vec{\beta} \in \mathbb{R}^{d-1}$. The implicit constant depends on C and Q.

Proof of theorem 5.8, case $d = 3$. Prescaling $\mathcal{P}$ by a constant amount we may suppose that Q is integral. We may also assume that the vector $(\alpha_1, \alpha_2) = R\vec{\beta}$ lies in $[0,1)\times[0,1)$, since $\mathcal{N}(R)$ is 1-periodic in these variables. Finally we assume $c > 4R^{-2}$ because both $\mathcal{N}(R)$ and $|\mathcal{P}|R^3$ are $O(R)$ when $c \ll R^{-1}$.

Writing $f(x,y) = (cR^2 - Q(x+\alpha_1, y+\alpha_2))/R$ we have

(5.10)
$$\frac{1}{2}\mathcal{N}(R) = \sum_{f(n_1,n_2)\geq 0}\sum \Big(\lfloor f(n_1,n_2)\rfloor + \frac{1}{2}\Big) = \sum_{f(n_1,n_2)\geq 0}\sum f(n_1,n_2) - \sum_{f(n_1,n_2)\geq 0}\sum \psi(f(n_1,n_2)).$$

Let χ the characteristic function of $Q(x+\alpha_1, y+\alpha_2) \leq cR^2$. Applying the Euler-Maclaurin formula firstly in n_2 and secondly in n_1, we have

$$\begin{aligned}\sum_{f(n_1,n_2)\geq 0}\sum f(n_1,n_2) &= \sum_{|n_1|\ll R\sqrt{c}} \left(\int \chi(n_1,y) f(n_1,y)\, dy + O(1)\right) \\ &= \int \chi(x,y) f(x,y)\, dydx + O(R)\end{aligned}$$

and the last integral is, of course, $\frac{1}{2}|\mathcal{P}|R^3$.

Using the Vaaler and Beurling polynomials of degree $M = \lfloor c^{1/2}R\rfloor$ (proposition 4.2), we get from (5.10)

$$\mathcal{E}(R) \ll \sum_{m=1}^{M} \frac{|S_m|}{m} + R^{1+\epsilon}$$

where

$$S_m = \sum_{Q(n_1,n_2)\leq M^2}\sum e\Big(\frac{m}{R}\Big(Q(n_1+\alpha_1, n_2+\alpha_2) - Q(\alpha_1,\alpha_2)\Big)\Big).$$

The summation domain has been changed at the cost of adding or removing at most $O(R)$ terms. Note this sum is exactly a truncated version of $\theta_{Q,\vec{v}}$ defined in (2.16), evaluated at m/R. Hence by corollary 2.20,

$$S_m \ll \frac{M^{2+\epsilon}}{q_m + M^2|q_m m/R - p_m|} \tag{5.11}$$

where the rational p_m/q_m is determined by the interval $\mathcal{A}_{p_m/q_m}$ of the Farey disection of order M where m/R lies. In particular, by proposition 1.3 it satisfies

$$\left|\frac{m}{R} - \frac{p_m}{q_m}\right| \leq \frac{1}{q_m(M+1)} \qquad \text{with} \quad q_m \leq M. \tag{5.12}$$

Let Ω be the set of all m in the interval $[1, M]$ for which $p_m \neq 0$, and note that for these m we have $q_m \asymp Rp_m/m$. Neglecting the term $M^2|q_m m/R - p_m|$ in (5.11),

$$\sum_{m\in\Omega}\frac{|S_m|}{m} \ll M^{2+\epsilon}R^{-1}\sum_{m\in\Omega}\frac{1}{p_m} = M^{2+\epsilon}R^{-1}\sum_{p\ll M^2R^{-1}}\frac{1}{p}\,\#\{m : p_m = p\}.$$

The last cardinality is $O(R^{1+\epsilon}/M)$ as $p_m = p$ and (5.12) imply that m must divide an integer in the interval $[Rp_m - R/M, Rp_m + R/M]$. This shows that the sum over Ω is $O(R^{1+\epsilon})$.

For the remaining terms $p_m = 0$ and $q_m = 1$, and (5.11) implies

$$\sum_{m\notin\Omega}\frac{|S_m|}{m} \ll RM^{\epsilon}\sum_{m\geq 1}\frac{1}{m^2} \ll R^{1+\epsilon}.$$

Hence, as claimed, $\mathcal{E}(R) \ll R^{1+\epsilon}$. □

PROOF OF THEOREM 5.8, CASE $d > 3$. Write $\vec{x} = (\vec{x}_1, \vec{x}_2)$ and $\vec{\beta} = (\vec{\beta}_1, \vec{\beta}_2)$ with $\vec{x}_1$, $\vec{\beta}_1 \in \mathbb{R}^2$ and $\vec{x}_2$, $\vec{\beta}_2 \in \mathbb{R}^{d-3}$. Let A be the matrix of Q and partition it as

$$A = \left(\begin{array}{c|c} A_1 & B \\ \hline B^t & A_2 \end{array}\right),$$

where $A_1 = (a_{ij})_{i,j=1}^2$ and $A_2 = (a_{ij})_{i,j=3}^{d-1}$. We have the identity $Q(\vec{x}) = Q_1(\vec{x}_1 + \vec{\gamma}) + Q_2(\vec{x}_2)$, where Q_1 (resp Q_2) is the positive definite quadratic form associated to A_1 (resp. $A_2 - B^t A_1 B$), and $\vec{\gamma} = A_1^{-1} B \vec{x}_2$. This is essentially "completing squares". Therefore, renaming $\vec{\gamma}$,

$$Q(\vec{x} + \vec{\beta}) = Q_1(\vec{x}_1 + \vec{\gamma}) + Q_2(\vec{x}_2 + \vec{\beta}_2). \tag{5.13}$$

Given $\vec{n}_2 \in \mathbb{Z}^{d-3}$, let us denote by $\mathcal{P}_{\vec{n}_2}$ the three-dimensional slice of $\mathcal{P}$ obtained by fixing $\vec{x}_2 = \vec{n}_2/R$, and by $\mathcal{N}_{\vec{n}_2}(R)$ the number of lattice points it contains after being dilated with scale factor R. By the three-dimensional case of this theorem and the decomposition (5.13),

$$\mathcal{N}(R) = \sum_{\vec{n}_2} \mathcal{N}_{\vec{n}_2}(R) = \sum_{\vec{n}_2} |\mathcal{P}_{\vec{n}_2}| R^3 + O(R^{d-2+\epsilon}),$$

both sums extended to the domain $Q_2(\vec{n}_2 + R\vec{\beta}_2) \leq cR^2$. A simple computation shows

$$|\mathcal{P}_{\vec{n}_2}| = \frac{\pi}{\sqrt{\det A_1}} \big(c - Q_2(\vec{n}_2/R + \vec{\beta}_2)\big)^2.$$

Applying the Euler-Maclaurin formula iteratively in one variable at a time we find

$$\frac{\pi}{\sqrt{\det A_1}} \sum_{\vec{n}_2} \big(c - Q_2(\vec{n}_2/R + \vec{\beta}_2)\big)^2 = \frac{\pi}{\sqrt{\det A_1}} \int \big(c - Q_2(\vec{x}_2/R)\big)^2 \, d\vec{x}_2 + O(R^{d-5})$$

and the main term in the right hand side is $|\mathcal{P}| R^{d-3}$. □

As in the last section, we are going to assume from now on that $c \in \mathbb{Q}$ and $\vec{\beta} = 0$ in order to obtain the Ω-results contained in theorem 5.2. The idea will be the same: showing that for arbitrarily large values of R, the number of points in the boundary of $R\mathcal{P}$, which will be denoted by $B(R)$, is $\Omega(R^{d-2}\eta(R))$, where η is the function defined in the statement of the theorem. Some reductions first: note that without loss of generality we may assume $c \in \mathbb{Z}$, and let $Q = \frac{a}{b} Q^*$ where Q^* is a primitive integral quadratic form. We also assume that $R \in \mathbb{Z}^+$, so that for each $\vec{n} \in \mathbb{Z}^{d-2}$ with $Q^*(\vec{n}) = Rn$ and $abn \leq cR$ we have that the lattice point $(b\vec{n}, cR - abn)$ is counted by $B(R)$. In other words,

$$B(R) \geq \sum_{n \leq \alpha R} r_{Q^*}(Rn) \qquad \text{with} \quad \alpha = \frac{c}{ab} \tag{5.14}$$

where $r_{Q^*}(k)$ is the number of representations of k by the quadratic form Q^*. For the remaining proofs we will not need to refer to Q anymore, and therefore we will write Q instead of Q^* for the sake of notational simplicity.

PROOF OF THEOREM 5.2, CASE $d = 3$. Let $r_1, r_2, \ldots, r_k$ be the solutions of

$$Q(r, 1) \equiv 0 \pmod{R}$$

and for each $1 \leq j \leq k$ and a fixed $0 < \delta < 1/2$ define

$$C_j = \{(x, y) \in \mathbb{Z}^2 \ : \ |y| \leq \delta R, \ |x| \leq \delta R, \ x \equiv r_j y \pmod{R}\}.$$

Choosing $\delta^2 < \lambda^{-1}\alpha$ with λ the greatest eigenvalue of the matrix of Q, we have that Q maps C_j into multiples of R less than αR^2. Hence the sum in (5.14) is at least $\#\bigcup_j C_j$. If we restrict y to $\gcd(y,R)=1$ then the sets C_j become disjoint, consequently

$$B(R) \geq k \min_j \#C_j - k\,\#\{y\in\mathbb{Z} \,:\, |y|<R,\ \gcd(y,R)>1\}. \tag{5.15}$$

For each fixed j, consider the remainders of $0r_j$, $1r_j$, $2r_j$,..., $\lfloor \delta R\rfloor r_j$ when divided by R. By the pigeonhole principle, if we subdivide $[0,R)$ into $\lceil\delta^{-1}\rceil$ equal subintervals, at least $\delta R/\lceil\delta^{-1}\rceil$ of the remainders lie in the same subinterval. In this way, we have at least $\delta R/\lceil\delta^{-1}\rceil$ pairs (u_ℓ, v_ℓ) such that $0\leq v_\ell\leq \delta R$ and all $u_\ell \equiv r_j v_\ell$ lie in the same subinterval of length $R/\lceil\delta^{-1}\rceil$. Hence $(u_\ell - u_1, v_\ell - v_1)\in C_j$ and it follows $\#C_j \geq \delta R/\lceil\delta^{-1}\rceil$. In this way, (5.15) assures

$$B(R) \geq k\frac{\delta^2 R}{1+\delta} + 2k(\phi(R)-R), \tag{5.16}$$

where ϕ stands for Euler's totient function. For large x, take R as the product of the primes $x\leq p\leq 2x$ such that $\left(\frac{\Delta}{p}\right)=1$ where Δ is the discriminant of Q. By the prime number theorem in arithmetic progressions, we have

$$\log R \sim \frac{x}{2} \qquad \text{and} \qquad \frac{\phi(R)}{R} = \prod_{p\mid R}(1-p^{-1}) = 1 + O\Big(\frac{1}{\log x}\Big). \tag{5.17}$$

The congruence $Q(r,1)\equiv 0$ admits two solutions modulo each of these primes p. Then by our choice of R we have that k equals 2 to the number of such primes that is at least $(\log R)/\log(2x)$. Substituting this and (5.17) in (5.16), we get the expected result. □

Proof of theorem 5.2, case $d=4$. Combining theorem 1 of [**8**] and theorem 2 of [**27**] we have

$$r_Q(n) = r_Q^{\mathrm{gen}}(n) + O(n^{13/28+\epsilon}) \qquad \text{for } n\notin\mathcal{S} \tag{5.18}$$

where $\mathcal{S}$ is a finite union of sets of the form $\{t_j m^2 \,:\, m\in\mathbb{Z}\}$ for some $t_j\in\mathbb{Z}$. Here r_Q^{gen} is the average number of representations by forms belonging to the same genus as Q that can be computed with Siegel mass formula (see §20.4 of [**62**] for the definitions and details). In lemma 6 of [**16**] this formula was written as

$$r_Q^{\mathrm{gen}}(n) = \frac{4\pi\sqrt{2n}}{\sqrt{D}} \sum_{d^2\mid n} d^{-1}U(n/d^2)L(1,\chi_{-2Dn/d^2}) \tag{5.19}$$

where D is the determinant of (the matrix associated to) Q, L is the L-function corresponding to the Kronecker symbol χ_m modulo $m=-2Dn/d^2$ and U is a certain $8D^2$-periodic function which is non-negative and not identically zero.

Assume $\gcd(R,2D)=1$ and for each $d^2\mid R$ choose n_d such that $U(n_d R/d^2)\neq 0$, then (5.18) and (5.19) together with (5.20) imply

$$B(R) \gg R\sum_{d^2\mid R} d^{-1}\mathcal{L}_d(R) + O(R^{27/14+\epsilon}) \tag{5.20}$$

where

$$\mathcal{L}_d(R) = \sum_{n\in\mathcal{A}} L(1,\chi_{-2DRn/d^2}) \qquad \text{with} \quad \mathcal{A} = \{n\asymp R \,:\, Rn\notin\mathcal{S},\ n\equiv n_d \pmod{8D^2}\}.$$

If $\mathcal{L}_d(R) \gg R$, choosing $R = \prod_{2D<p\le x} p^2$ we have $\log R \sim 2x$ and

$$B(R) \gg R^2 \prod_{2D<p\le x} (1+p^{-1}) + O(R^{27/14+\epsilon}) \gg R^2 \log\log R.$$

It remains to prove $\mathcal{L}_d(R) \gg R$. Expanding the L-functions, we can write $\mathcal{L}_d(R)$ as

$$S_1+S_2+S_3 := \sum_{m_1} \frac{1}{m_1} \sum_{n\in\mathcal{A}} \chi_{d_n}(m_1) + \sum_{m_2} \frac{\chi_{-2DR'}(m_2)}{m_2} \sum_{n\in\mathcal{A}} \chi_n(m_2) + \sum_{n\in\mathcal{A}} \sum_{m_3} \frac{\chi_{d_n}(m_3)}{m_3}$$

where $d_n = -2DR'n$, $R' = R/d^2$, m_1 runs over the squares in $[1, R^{1+\epsilon}]$, m_2 over the non-squares coprime to $2DR'$ in the same interval and $m_3 > R^{1+\epsilon}$. Trivially, $S_1 \gg R$. By Pólya-Vinogradov inequality $S_3 \ll \sum_{n\in\mathcal{A}} R^{-\epsilon} \ll R^{1-\epsilon}$. There are $O(R^{1/2})$ values of $n \asymp R$ with $Rn \in \mathcal{S}$ that when added to $\mathcal{A}$ give a negligible contribution $O(R^{1/2}\log R)$ to S_2, and hence we can drop the condition $Rn \notin \mathcal{S}$ in S_2. On the other hand, the congruence condition $n \equiv n_d$ can be detected inserting $\sum_\chi \chi(n)\overline{\chi}(n_d)/\phi(8D^2)$ where χ runs over the characters modulo $8D^2$. Since $\gcd(m_2, 2DR') = 1$, the product $\chi(n)\chi_n(m_2)$ as a function of n is a non-principal character modulo $8D^2m_2$ and Pólya-Vinogradov inequality proves $S_2 \ll R^{1/2+\epsilon}$. Therefore, simply bounding below S_1 by the summand corresponding to $m_1 = 1$, we conclude $\mathcal{L}_d(R) \sim S_1 \gg R$. □

Proof of theorem 5.2, case $d \ge 5$. For $d \ge 6$ we have by corollary 11.3 of [**61**] the estimate $r_Q(m) \asymp m^{(d-3)/2}$ as long as m is sufficiently large and $Q(\vec{x}) \equiv m$ is solvable modulo 2^7D^3 with D the determinant of Q. Taking $m = Rn$ with R a large multiple of 2^7D^3, both conditions are fulfilled and the result follows from (5.14).

If $d = 5$, corollary 11.3 of [**61**] gives for $2^7D^3 \mid R$

$$B(R) \gg R \sum_{n\le \alpha R} n \prod_{p\mid Rn} (1+\chi_D(p)p^{-1}) \qquad \text{with} \quad \chi_D(p) = \left(\frac{D}{p}\right). \tag{5.21}$$

Let P_D the product of the primes $p \le x$ such that $\chi_D(p) = 1$. By the prime number theorem in arithmetic progressions, we have

$$\log P_D \sim \frac{x}{2} \qquad \text{and} \qquad \prod_{p\mid P_D} (1+p^{-1}) \gg \sqrt{\log x} \sim \sqrt{\log\log P_D}.$$

Choosing $R = 2^7D^3P_D$ in (5.21), we have

$$B(R) \gg R \prod_{p\mid P_D} (1+p^{-1}) \cdot \sum_{n\le \alpha R} n \prod_{p\mid n} (1-p^{-1}).$$

The sum equals that of $\phi(n)$, that is comparable to R^2 (theorem 330 of [**46**]). □

CHAPTER 6

Lattice points in revolution bodies

This chapter focuses in the results contained in the article "Lattice points in revolution bodies (II)" [**21**], joint work with F. Chamizo.

6.1. Main results

In §4.1 we saw the huge difference between the known results for lattice point counting problems associated to general smooth convex bodies and to balls. In particular, in three dimensions, the best known upper bound for α_K in the former case is Guo's $\alpha_K \leq 231/158 \approx 1.462$, while for the unit ball we have Heath-Brown's $\alpha_B \leq 21/16 = 1.3125$. The improvement, slightly below 0.15, is usually a huge gap when dealing with cancellation of exponential sums, and is made possible only thanks to the arithmetic of quadratic forms.[1]

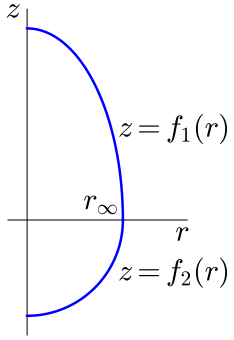
F. Chamizo noted in [**15**] that if one assumes rotational symmetry around a coordinate axis then one can obtain intermediate results even from the simplest van der Corput's estimates. He considered three-dimensional smooth convex bodies of the form

$$K = \{(x,y,z) \in \mathbb{R}^3 \,:\, f_2(r) \leq z \leq f_1(r),\ 0 \leq r \leq r_\infty\} \tag{6.1}$$

where $r = \sqrt{x^2+y^2}$.

In other words, K is the solid generated by the rotation around the z-axis of the curve

$$\gamma(t) = \begin{cases} (t, 0, f_1(t)) & 0 \leq t \leq r_\infty \\ (2r_\infty - t, 0, f_2(2r_\infty - t)) & r_\infty \leq t \leq 2r_\infty \end{cases}$$



Theorem 1.1 of [**15**] reads:

THEOREM. *Let $K \subset \mathbb{R}^3$ be a smooth convex body which is also body of revolution, and suppose that the functions $\frac{1}{r} f_i'''(r)$ (extended by continuity to $r = 0$) do not vanish for $0 \leq r < r_\infty$, where $i = 1, 2$. Then the inequality $\alpha_K \leq 11/8$ holds.*

[1] To put this into context, note that the last improvement of Gauss' circle problem goes from Huxley's $131/208 \approx 0.62981$ to Bourgain and Watt's $517/824 \approx 0.62743$, barely 0.00238 of difference.

Note that $11/8 = 1.375$ is actually closer to Heath-Brown's than to Guo's result. Nevertheless Chamizo had to assume the very unnatural hypothesis concerning the nonvanishing of the third derivative of the generatrix function. The reason for this is that the exponential sum was estimated via an application of a Weyl step followed by van der Corput's lemma, and hence one must control that the size of a third derivative of the phase is neither too big nor too small (*cf.* §4.4). Although this third derivative is in principle mixed —the Weyl step and the van der Corput estimation happening in different variables— these two variables become interlinked by the rotational symmetry of the convex body. It is in fact in this way that the two variables are "glued" together into only one variable running over a longer interval, allowing for greater power savings from the application of the van der Corput method. At the end of the day the mixed third derivative of the phase function can actually be seen to correspond to the third derivative of the generatrix.

When this kind of conditions involving derivatives of the phase function come into play it is usually a defect of the method used to estimate the exponential sum. If the third derivative becomes too small, since each derivative is usually smaller than the last, it means that the affected portion of the exponential sum must be treated with methods which involve derivatives of lesser degree. In practice, however, things are often not that simple, and this kind of conditions have been historically a hassle, even when obtaining results for the circle and divisor methods (see [**55**] and [**70**]). The advent of the discrete Hardy-Littlewood method [**57, 58**] has more or less resolved this issue for $d = 2$, but the problem still persists when $d > 2$. In fact, most of the technical part of Guo's paper [**39**] revolves around showing that some combinations of partial derivatives never vanish all of them at the same time.

In the article [**21**] we did not succeed at removing the nonvanishing condition, but we were able to replace it with the following much weaker version:

THEOREM 6.1. *Let $K \subset \mathbb{R}^3$ be a smooth convex body of revolution, and suppose that the third derivative of the generatrix functions f_i''' only have zeros of finite order for $0 \leq r < r_\infty$, where $i = 1, 2$. Then the inequality $\alpha_K \leq 11/8$ holds.*

By zeros of finite order we mean that $f_i'''(r) = 0$ implies we can find an integer $n > 3$ such that $f_i^{(n)}(r) \neq 0$. In particular, this is satisfied whenever the boundary of K is real-analytic. The result also holds if in the definition (6.1) we take $r = \sqrt{Q(x - \alpha, y - \beta)}$ with Q a positive definite rational quadratic form and $\alpha, \beta \in \mathbb{R}$. In other words, theorem 6.1 extends to the case in which the horizontal sections are rational ellipses with a common center when projected onto the xy-plane.

The idea of the proof is the following: we transform the problem via Poisson summation into estimating an exponential sum, as it is customary; and then slice the sum diadically in pieces corresponding to the zeros of $f_i'''(r)$. For the pieces where van der Corput's lemma falls short the phase is almost linear, and we are in position to apply the Kuzmin-Landau lemma. This, by itself, is not good enough, as the derivative of the phase function might happen to be close to an integer way too often. Showing that this cannot be the case requires —in some ranges— studying certain Diophantine properties of a Taylor coefficient of the phase function. This goes beyond the utterly analytic treatment in the classical (van der Corput's) theory of exponential sums and vaguely resembles to the situation in [**10**] (the seminal paper for the discrete Hardy-Littlewood method) in which the arithmetic properties of the Taylor coefficients play a fundamental role.

While working on this problem, the first obvious step was to look into those examples which seem the most pathological, and in this case the nonvanishing hypothesis is blatantly violated when both functions f_i are second order polynomials, *i.e.* K is a revolution paraboloid (or, more generally, an elliptic paraboloid). This is precisely the problem treated in chapter 5, for which the conjecture was obtained in the rational case thanks to the automorphic properties of the exponential sum. In some sense a related phenomenon is happening here, as very close to a zero of $\frac{1}{r}f_i'''(r)$ the function $f_i(r)$ essentially looks like a parabola, and some of the arithmetic leaks in in the form of the aforementioned Diophantine properties of the Taylor coefficient. Since we are only aiming for the exponent 11/8 (as we cannot do better anyway because most of the boundary of K cannot be well approximated by parabolas), an adapted version using the van der Corput method is enough and we can skip modular forms altogether.

Since we are only involving derivatives up to order three, theorem 6.1 should remain true if K is of class $\mathcal{C}^3$ and the zeros of f_i''' are isolated and f_i''' decays as a fixed power of the distance to the closest zero. When the zeros are dense or of "infinite order" the method fails because one has to chop the exponential sum into too many pieces, most of them too small to have appreciable cancellation. This, together with the fact that in the most extreme case one obtains not only the 11/8 but the full conjecture, leads me to believe that the remaining hypothesis is still an artifact of the methods used, and the exponent 11/8 should hold for any revolution convex body of this regularity. Sadly, this is probably out of reach with the existing methods.

6.2. The exponential sum

Our starting point is the truncated Hardy-Voronoï formula provided by proposition 4.1, which for convenience we copy here:

$$\mathcal{E}(R) = -\frac{R'}{\pi} \sum_{\vec{0}\neq\vec{n}\in\mathbb{Z}^3} \eta(\delta\|\vec{n}\|) \frac{\cos\left(2\pi R' g(\vec{n})\right)}{\|\vec{n}\|^2\sqrt{\kappa(\vec{n})}} + O(R^{2+\epsilon}\delta). \tag{6.2}$$

In this formula η is a certain even non-negative smooth function compactly supported in $[-1,1]$, $\delta = R^{-c}$ for some fixed $0 < c < 2$, R' depends on R in a non-explicit way but always stays at a fixed distance from it, g is defined by $g(\vec{n}) = \sup\{\vec{x}\cdot\vec{n} \,:\, \vec{x}\in K\}$ and $\kappa(\vec{n})$ stands for the Gaussian curvature of the boundary of K at the point whose unit outer normal is $\vec{n}/\|\vec{n}\|$.

As we commented in chapter 4, the larger c is chosen the smaller the error term is, but also the longer the exponential sum becomes, and since the van der Corput method provides power savings on the length of the sum, the larger the corresponding bound will be. Usually one leaves c as an unknown, works out all the details and then chooses the value of c which balances both error terms. Once this is done one can either write the article this way, or directly fix c to the value that magically makes all extra error terms vanish. Chamizo's original article [**15**] is written in the first way, which is a good starting point for the reader who wants to know where the 11/8 comes from and why it cannot be improved using these techniques. Here, however, since the proof is substantially more convoluted, it results more convenient to directly fix $c = 5/8$.

When K is a body of revolution body all the functions of $\vec{n}$ involved in the expression (6.2) for $\mathcal{E}$ are invariant under rotations on the first two variables. Writing

$\vec{n} = (n_1, n_2, m)$ and $n = n_1^2 + n_2^2$, and grouping the terms with common n,

(6.3)
$$\mathcal{E}(R) = -\frac{R'}{\pi} \Re \sum_{n>0} \sum_{0 \neq m \in \mathbb{Z}} r_2(n) \eta\big(\delta\sqrt{n+m^2}\big) \frac{e\big(R'h(n,m)\big)}{(n+m^2)\sqrt{\kappa_1(n,m)}} + O\big(R^{11/8+\epsilon}\big)$$

where $h(n,m) = g(\sqrt{n}, 0, m)$ and $\kappa_1(n,m) = \kappa(\sqrt{n}, 0, m)$. Note we have estimated trivially the terms corresponding to $nm = 0$.

Before continuing we are going to take a moment to examine the case when K is not a revolution body with respect to the z-axis, but it is defined in terms of $r = \sqrt{Q(x-\alpha, y-\beta)}$ for Q a positive definite quadratic form with rational coefficients. Note we can always write $K = B^{-1}K' + \vec{\tau}$, where K' is a revolution body, B is a 3×3 matrix whose top-left 2×2 block B_1 satisfies $\vec{x}^t B_1^t B_1 \vec{x} = Q(\vec{x})$ and the rest of the matrix coincides with the identity, and $\vec{\tau} = (\alpha, \beta, 0)^t$. A simple computation shows $g(\vec{n}) = g'(B^{-t}\vec{n}) + \vec{\tau} \cdot \vec{n}$ where g' is the function associated to K'. To take advantage of the invariance of g' we must group the terms of the sum according to $n = Q^*(n_1, n_2)$, where $Q^*(\vec{x}) = \vec{x}^t B_1^{-1} B_1^{-t} \vec{x}$ is the dual form of Q, whose associated matrix is the inverse of that of Q. Indeed,

$$e\big(R'g(\vec{n})\big) = e\big(R'h'(n,m)\big) \cdot e\big(R'(\alpha n_1 + \beta n_2)\big) \qquad \text{if} \quad Q^*(n_1, n_2) = n,$$

with $h'(n,m) = g'(\sqrt{n}, 0, m)$. Also without loss of generality, prescaling K, we may assume Q^* has integer coefficients and hence n runs over the integers.

Conveniently, we also have

$$\|\vec{n}\|^2 \sqrt{\kappa(\vec{n})} = |\det B| \, \|B^{-t}\vec{n}\|^2 \sqrt{\kappa'(B^{-t}\vec{n})}.$$

This can be shown directly from the definition of Gaussian curvature,[2] but a shortcut is to use the properties of the Fourier transform, together with the expression we have for g in terms of g', to see that the substitution is possible in the expansion (4.1), and then follow again the steps of the proof of proposition 4.1. In any case, to fully exploit the geometry of the problem at hand we must go back to this proof anyway, and check that the function $\eta(\delta\|\vec{n}\|)$ may be replaced by $\eta(\delta\|B^{-t}\vec{n}\|)$ under the same hypotheses.

With these modifications we can now carry out the argument above, grouping the terms corresponding to the same value of $n = Q^*(n_1, n_2)$ together, and recover (6.3) with r_2 replaced by $r_{Q^*,\vec{v}}$ (defined in (2.15)) for $\vec{v} = R'(\alpha, \beta)^t$, the functions h' and κ' corresponding to the revolution body K' and an extra factor $|\det B|^{-1}$. For the sake of simplicity we will denote $r_{Q^*,\vec{v}}$ by r_2^* and drop the prime on h and κ. The upper bound $r_2^*(n) \ll n^\epsilon$ holds in general by virtue of (**I**.16), justifying that we can neglect the terms with $nm = 0$.

The next step is to sum by parts in (6.3) to remove all factors but the arithmetic function r_2^* and the exponential. It is important this is done after grouping the terms, as if we had summed by parts directly in (6.2) the resulting exponential sum would have been supported in rectangular boxes and grouping the terms in circles (or ellipses) would result impossible.

[2] In general, if M is any invertible 3×3 matrix, the Gaussian curvature of MK is related to that of K by the formula

$$\kappa_{MK}(\vec{n}) \cdot \|\vec{n}\|^4 (\det M)^2 = \kappa_K\big(M^t\vec{n}\big) \cdot \|M^t\vec{n}\|^4.$$

This is stated without proof in [**15**].

To sum by parts it is best to first divide the sum (6.3) into two halves, depending on the sign of m. In fact, it suffices to estimate the half S^+ corresponding to $m > 0$ (which, as we will see below, arises from the north half of K delimited by f_1). Indeed, by the properties of the Fourier transform, the sum corresponding to the specular reflection of K through the plane $z = 0$ is exactly the same sum, but with the sign of m reversed. Therefore if we succeed at estimating the half S^+ for every K, the same argument applied to its specular reflection yields the same bound for the other half.

Summing by parts in two variables in this case is particularly easy because all boundary terms vanish as η is compactly supported (see the appendix). Hence, writing the main term in integral form,

$$S^+ = -\frac{R'}{\pi|\det B|}\Re \iint \sum_{n\le u}\sum_{m\le v} r_2^*(n)e\big(R'h(n,m)\big)\frac{\partial^2}{\partial u\partial v}\frac{\eta\big(\delta\sqrt{u+v^2}\big)}{(u+v^2)\sqrt{\kappa_1(u,v)}}\,du\,dv,$$

the integral extended over the rectangle $[1,\delta^{-2}]\times[1,\delta^{-1}]$. Multiplying and dividing the integrand by $u+v^2$ and estimating trivially,

$$S^+ \ll \sup_{N,M^2\le\delta^{-2}} \frac{R^{1+\epsilon}}{N+M^2}\left|\sum_{1\le n\le N}\sum_{1\le m\le M} r_2^*(n)e\big(R'h(n,m)\big)\right|, \tag{6.4}$$

as long as we can guarantee

$$\iint (u+v^2)\left|\frac{\partial^2}{\partial u\partial v}\frac{\eta\big(\delta\sqrt{u+v^2}\big)}{(u+v^2)\sqrt{\kappa_1(u,v)}}\right|du\,dv \ll R^\epsilon.$$

After performing the change of variables $u\mapsto u^2$ and changing to polar coordinates (note $\kappa_1(u^2,v)=\kappa(u,0,v)$ depends smoothly on the angle $\theta=\arctan v/u$), the integral becomes

$$\iint \rho^3\left|\frac{\partial^2}{\partial u\partial v}\frac{\eta(\delta\rho)}{\rho^2\sqrt{\kappa(\theta)}}\right|d\rho\,d\theta.$$

The operator $\frac{\partial^2}{\partial u\partial v}$ decomposes as sum of differential operators in polar coordinates which, neglecting the dependence on θ, are of the form $\frac{\partial^2}{\partial\rho^2}$, $\rho^{-1}\frac{\partial}{\partial\rho}$ or ρ^{-2}. This shows the integral is bounded by $\int_1^{\delta^{-1}}\big(\rho^{-1}+\delta+\delta^2\rho\big)\,d\rho \ll \log R$.

Now that (6.4) is established we can cut the sum dyadically to prepare it for the van der Corput method. Actually, what we really want is to cut the sum dyadically in the image of the partial derivative of h involved in the van der Corput estimation, to make sure we control its size in each piece. If the zeros and poles of this function are of finite order, this amounts to splitting the domain of the sum dyadically around these points. In our case we will see in §6.4 that the size of the appropriate partial derivative essentially depends on $|f_1'''(r)|$ for $r=(f_1')^{-1}(\sqrt{n}/m)$, and hence the sum should be split dyadically as n/m^2 approaches the squares of slopes of f_1 at the points where f_1''' either vanishes or diverges.

Since we are only dealing with half of the sum S^+, the contour function f_2 will not make any appearance in the rest of the chapter, and therefore we can conveniently rename f_1 to f. Let $0=r_0<r_1<\cdots<r_{j_0-1}$ be the zeros of f''' in $[0,r_\infty)$, which are necessarily finite in number as they are of finite order and f''' diverges as $r\to r_\infty$, and fix any r_{j_0} satisfying $r_{j_0-1}<r_{j_0}<r_\infty$. Denote $u_j=(f'(r_j))^2$ for $0\le j\le j_0$. We split the summation domain of (6.4) dyadically in

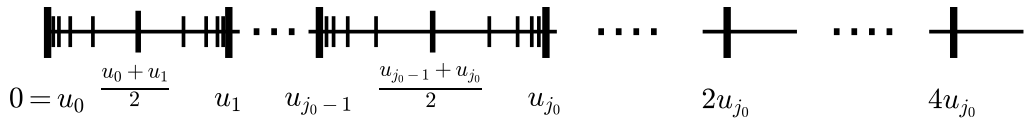


FIGURE 6.1. The dyadic decomposition of the sum S^+.

m as $m \to \infty$, and in n/m^2 as it approaches either some u_j or ∞ (see figure 6.1), obtaining smaller sums of the form

$$S(U_1, U_2, M) = \sum_{\substack{U_1 \le n/m^2 < U_2 \\ M \le m < 2M \\ 1 \le n \le N}} \sum r_2^*(n) e\big(Rh(n,m)\big). \tag{6.5}$$

For simplicity we have also renamed R' to R. The dependence in N will not bear any importance in the rest of the proof and for the sake of clarity we make it implicit.

The trivial estimate $S(U_1, U_2, M) \ll R^\epsilon (U_2 - U_1) M^3 + R^\epsilon M$ shows we can neglect all the pieces sufficiently close to each u_j, leaving at most $O(\log R)$ pieces to estimate. Theorem 6.1 therefore follows from the following two theorems. The part $0 \le n/m^2 < u_{j_0}$ of the double sum in (6.4) is covered by theorem 6.2, while the part $n/m^2 \ge u_{j_0}$ is covered by theorem 6.3 (note $U \le N/M^2$ or the sum is empty).

THEOREM 6.2. *Given $\epsilon > 0$ and $0 \le j < j_0$, for any $R > 1$, $2M \le R^{5/8}$ and $0 < U \le (u_{j+1} - u_j)/4$ we have*

$$\big|S(u_{j+1} - 2U, u_{j+1} - U, M)\big| + \big|S(u_j + U, u_j + 2U, M)\big| \ll M^2 R^{3/8+\epsilon}.$$

THEOREM 6.3. *Given $\epsilon > 0$, for any $R > 1$, $2M \le R^{5/8}$ and $u_{j_0} \le U \le R^{5/4} M^{-2}$ we have*

$$S(U, 2U, M) \ll U M^2 R^{3/8+\epsilon}.$$

6.3. Weyl step

In order to be able to estimate the sum S given by (6.5) using the van der Corput method we must first get rid of the arithmetic function r_2. We do this by performing a Weyl step (*cf.* §4.4).

PROPOSITION 6.4. *Let S as before and fix $\epsilon > 0$. For any $1 \le M \le R^{5/8}$, $0 < U_1 < U_2 \le R^{5/4}$ and $1 \le L \ll M$, satisfying $U_2 - U_1 = U$ and $U_2 L + 1 \ll UM$, we have*

$$\big|S(U_1, U_2, M)\big|^2 \ll R^\epsilon (U^2 M^6 L^{-1} + U M^3 T)$$

where $T = T(U_1, U_2, M, L)$ is given by

$$T = \frac{1}{L} \sum_{1 \le \ell \le L} \left| \sum_{\substack{U_1 \le n/(m+\ell)^2, n/m^2 < U_2 \\ M \le m, m+\ell < 2M \\ 1 \le n \le N}} \sum e\Big(R\big(h(n, m+\ell) - h(n,m)\big)\Big) \right|. \tag{6.6}$$

PROOF. Consider

$$\psi_{n,m} = \begin{cases} e(Rh(n,m)) & \text{if } U_1 \le n/m^2 < U_2,\ M \le m < 2M \text{ and } n \le N, \\ 0 & \text{otherwise.} \end{cases}$$

It suffices to prove the inequality when L is an integer. We may therefore write

$$LS = \sum_{M-L\le m<2M} \sum_{U_1 m^2\le n<U_2(m+L)^2} r_2^*(n) \sum_{1\le \ell\le L} \psi_{n,m+\ell}.$$

The length of the first sum is $\ll M$ and the length of the second one $\ll UM^2$, hence squaring and applying Cauchy-Schwarz,

$$L^2S^2 \ll R^\epsilon UM^3 \sum_{M-L\le m<2M} \sum_{U_1 m^2\le n<U_2(m+L)^2} \sum_{1\le \ell_1,\ell_2\le L} \psi_{n,m+\ell_1}\overline{\psi_{n,m+\ell_2}}.$$

Separating the diagonal contribution $\ell_1 = \ell_2$ and interchanging the summation order, which can be done because $\psi_{n,m}$ keeps track of the summation domain,

$$L^2S^2 \ll R^\epsilon U^2M^6L + R^\epsilon UM^3 \Re \sum_{1\le \ell_2<\ell_1\le L} \sum_n \sum_m \psi_{n,m+\ell_1}\overline{\psi_{n,m+\ell_2}}.$$

To obtain the desired inequality it is enough to perform the change of variables $m \mapsto m-\ell_2$ and group the terms corresponding to each value of $\ell = \ell_1 - \ell_2$. □

6.4. The function h

In this section we prove the estimates we need about the function h. Note that the convexity of $-f$ implies that $-f' : [0, r_\infty) \to \mathbb{R}^+$ is one-to-one, and therefore its inverse function ϕ is well-defined.

LEMMA 6.5. *We have the identity*

$$\frac{\partial}{\partial m} h(n,m) = F\big(n/m^2\big) \quad \textit{where} \quad F(u) = f\big(\phi(\sqrt{u})\big).$$

PROOF. The supremum $\sup\{\vec{x}\cdot\vec{n} \,:\, \vec{x}\in K\}$ defining $g(\vec{n})$ is clearly attained on a point $\vec{x}_0$ on the boundary of K which is a critical point for the function $\vec{x}\cdot\vec{n}$. By geometrical considerations this can only happen if the tangent plane at $\vec{x}_0$ is orthogonal to $\vec{n}$, leaving only two possibilities for $\vec{x}_0$. The supremum is therefore always attained on the point whose unit outer normal vector coincides with $\vec{n}/\|\vec{n}\|$.

After the previous considerations, the function $h(n,m) = g\big(\sqrt{n},0,m\big)$ for $m>0$ must be univocally determined by

$$h(n,m) = r\sqrt{n} + f(r)m \qquad \text{where} \qquad f'(r) = -\frac{\sqrt{n}}{m}.$$

Differentiating implicitly with respect to m,

$$\frac{\partial}{\partial m} h(n,m) = \frac{\sqrt{n}}{mf''(r)}\left(\frac{\sqrt{n}}{m} + f'(r)\right) + f(r).$$

The first term vanishes, while $f(r) = F\big(n/m^2\big)$ as desired. □

The estimates for h near the "bad" points u_j will depend on the order of vanishing of $f'''(r)$. By definition, each u_j is the preimage by the function $\phi(\sqrt{u})$ of a zero r_j of $f'''(r)$, except the last one which is added for convenience. If $r_j \neq 0$ we define d_j as the unique non-negative integer satisfying $f'''(r) \asymp (r-r_j)^{d_j}$ as $r\to r_j$. For $r_0 = 0$ we define d_0 as the unique non-negative integer satisfying $f'''(r) \asymp r^{2d_0+1}$ as $r\to 0^+$. We also set $d_\infty = -5/2$.

LEMMA 6.6. *We have $F'(u) \asymp (1+u)^{-3/2}$ for $0\le u<\infty$. We also have $F''(u)\neq 0$ for $u\neq u_j$, $0\le j\le j_0$, and*

$$F''(u) \asymp (u-u_j)^{d_j} \quad \textit{as} \quad u\to u_j \qquad \textit{and} \qquad F''(u) \asymp u^{d_\infty} \quad \textit{as} \quad u\to\infty.$$

PROOF. Let $k(r)$ denote the curvature of $r \mapsto \big(r, f(r)\big)$, which admits the explicit formula

$$f''(r) = k(r)\big(1 + |f'(r)|^2\big)^{3/2}, \tag{6.7}$$

and set $c(u) = k(\phi(\sqrt{u}))$. Differentiating F and recalling that ϕ is the inverse function of $-f'$ we have $F'(u) = \big[2f''(\phi(\sqrt{u}))\big]^{-1}$. Differentiating again and using (6.7) in the form $f''(\phi(\sqrt{u})) = c(u)(1+u)^{3/2}$ we obtain

$$F'(u) = \frac{1}{2c(u)(1+u)^{3/2}},$$
$$F''(u) = \frac{f'''(\phi(\sqrt{u}))}{4\big(c(u)\big)^3(1+u)^{9/2}u^{1/2}}.$$

Now all but the last claim of the lemma is clear as $c(u) \asymp 1$ and $\phi(\sqrt{u})$ has nonvanishing derivative for $u > 0$, and behaves like $C\sqrt{u}$ for some $C \neq 0$ as $u \to 0^+$. To establish the last claim, we note that by (6.7) and L'Hôpital's rule,

$$k(r_\infty) = \lim_{r\to r_\infty} \frac{f''(r)}{|f'(r)|^3} = \lim_{r\to r_\infty} \frac{-f'''(r)}{3\big(f'(r)\big)^2 f''(r)} = \lim_{u\to\infty} \frac{-f'''(\phi(\sqrt{u}))}{3c(u)(1+u)^{3/2}u}.$$

Therefore $f'''(\phi(\sqrt{u})) \asymp u^{5/2}$ when $u \to \infty$, and $F''(u) \asymp u^{-5/2}$. □

We use lemma 6.6 to give estimates for some derivatives of h.

PROPOSITION 6.7. *Let $(n,m) \in \big(\mathbb{R}^+\big)^2$ with $m \asymp M$. If $n/m^2 < u_{j_0}$ let U be distance of n/m^2 to the closest u_i, say u_j. If $n/m^2 \geq u_{j_0}$ take $U = n/m^2$ and $j = \infty$. Then*

$$\frac{\partial^3 h}{\partial n^2 \partial m}(n,m) \asymp \frac{U^{d_j}}{M^4}.$$

PROOF. By lemma 6.5 the partial derivative is $m^{-4}F''(n/m^2)$ and the result follows from lemma 6.6. □

PROPOSITION 6.8. *Let $(n,m) \in \big(\mathbb{R}^+\big)^2$ with $m \asymp M$ and fix j with $d_j > 0$. If $U = |n/m^2 - u_j|$ is small enough, then*

$$\frac{\partial^3 h}{\partial n \partial m^2}(n,m) \asymp \frac{1}{M^3}.$$

PROOF. The partial derivative here is $-2m^{-3}\big(F''(n/m^2)n/m^2 + F'(n/m^2)\big)$. By lemma 6.6 the function F' remains positive and bounded in bounded subintervals of $\mathbb{R}^+$, while $F''(n/m^2)n/m^2 \to 0$ when $U \to 0$. □

PROPOSITION 6.9. *Let $(n,m) \in \big(\mathbb{R}^+\big)^2$ with $m \asymp M$ and fix j with $d_j > 0$. If $U = |n/m^2 - u_j|$ is small enough and $1 \leq \ell \leq UM$,*

$$\frac{\partial h}{\partial n}(n, m+\ell) - \frac{\partial h}{\partial n}(n,m) = C_j \frac{\ell}{m(m+\ell)} + O\left(\frac{\ell U^{d_j+1}}{M^2}\right)$$

for some constant $C_j \neq 0$.

PROOF. We express the left hand side as

$$\int_0^\ell \frac{\partial^2 h}{\partial n \partial m}(n, m+t)\, dt = \int_0^\ell F'\left(\frac{n}{(m+t)^2}\right)\frac{dt}{(m+t)^2}$$

$$= \int_0^{\ell} \left[F'(u_j) + \int_{u_j}^{n/(m+t)^2} F''(v)\, dv \right] \frac{dt}{(m+t)^2}$$
$$= F'(u_j) \frac{\ell}{m(m+\ell)} + O\left(\frac{\ell U^{d_j+1}}{M^2} \right).$$

To bound the error term we have applied lemma 6.6 noting that $n/(m+t)^2 - u_j = O(U)$ for $0 \leq t \leq UM$. □

6.5. The van der Corput estimate

In this section we use the estimates we have just obtained, together with the van der Corput lemma, to estimate the sum (6.5) in certain ranges, covering part of theorems 6.2 and 6.3. Up to here there is nothing essentially new in comparison with Chamizo's original article [**15**], and in fact if $d_j = 0$ for all $1 \leq j \leq j_0$ we readily recover the original version of theorem 6.1.

To simplify the proofs, we will assume from now on that $UM \geq R^{3/8}$, as otherwise the trivial estimate $S \ll R^{\epsilon} UM^3 + R^{\epsilon} M$ suffices to prove the desired inequalities. We will also refer to the arguments of S in the statements of theorems 6.2 and 6.3 as U_1 and U_2 for the sake of convenience.

Proposition 6.10. *Let R, M, U, U_1 and U_2 be as in the hypotheses of either theorem 6.2 or 6.3, setting $j = \infty$ in the second case. Then*

$$\sum_n e\big(R(h(n, m+\ell) - h(n,m))\big) \ll R^{1/2} \ell^{1/2} U^{(d_j+2)/2} + R^{-1/2} \ell^{-1/2} U^{-d_j/2} M^2,$$

where the range of the summation is $U_1(m+\ell)^2 \leq n < \min\big(U_2 m^2, N\big)$ for $m \asymp M$.

Proof. By the mean value theorem and proposition 6.7 we have

$$\frac{\partial^2}{\partial n^2}\big(h(n, m+\ell) - h(n,m)\big) = \ell \frac{\partial^3 h}{\partial n^2 \partial m}(n, \tilde{m}) \asymp \ell \frac{U^{d_j}}{M^4}.$$

Applying now the van der Corput lemma (proposition 4.4),

$$\sum_n e\big(R(h(n, m+\ell) - h(n,m))\big) \ll UM^2 \big(R\ell U^{d_j} M^{-4}\big)^{1/2} + \big(R\ell U^{d_j} M^{-4}\big)^{-1/2}.$$

This concludes the proof. □

Proposition 6.11. *Theorem 6.2 holds when $d_j = 0$, 1, or when $d_j \geq 2$ and $U \gg R^{-5/(8d_j - 8)}$.*

Proof. Note that since $U_2 \ll 1$ we are in position to apply proposition 6.4 as long as we take $L \leq UM$. Using proposition 6.10 to bound $T(U_1, U_2, M, L)$ we obtain

$$R^{-\epsilon} M^{-4} |S|^2 \ll L^{-1} U^2 M^2 + R^{1/2} L^{1/2} U^{(d_j+4)/2} + R^{-1/2} L^{-1/2} U^{(2-d_j)/2} M^2. \tag{6.8}$$

We choose $L = \min\big(R^{1/2} U^{-d_j}, UM\big)$. If $L = R^{1/2} U^{-d_j}$ then using $M \leq R^{5/8}$ we obtain $M^{-4}|S|^2 \ll R^{3/4+\epsilon}$, as desired. Hence assume $L = UM$ and $U^{d_j+1} < R^{1/2} M^{-1}$. We have

$$R^{-\epsilon} M^{-4} |S|^2 \ll UM + R^{1/2} U^{(d_j+5)/2} M^{1/2} + R^{-1/2} U^{(1-d_j)/2} M^{3/2}.$$

Using the inequality $U^{d_j+1} < R^{1/2} M^{-1}$ on the second summand and the hypotheses of this proposition we conclude again $M^{-4}|S|^2 \ll R^{3/4+\epsilon}$. □

PROOF OF THEOREM 6.3. We proceed similarly as in the previous proof. Note that now $U_2 \asymp U$ and we may take $1 \le L \le M$ in proposition 6.4. Using proposition 6.10 to bound $T(U_1, U_2, M, L)$ we obtain exactly the same bound (6.8) with $d_\infty = -5/2$:

$$R^{-\epsilon} M^{-4} |S|^2 \ll L^{-1} U^2 M^2 + R^{1/2} L^{1/2} U^{3/4} + R^{-1/2} L^{-1/2} U^{9/4} M^2.$$

The choice $L = \min(R^{1/2}, M)$ also works in exactly the same way, using $U \le R^{5/4} M^{-2}$ and $M \le R^{5/8}$ (or $M \le R^{1/2}$ if $L = M$), to show $M^{-4}|S|^2 \ll U^2 R^{3/4+\epsilon}$. □

6.6. Diophantine approximation of the phase

As U gets smaller than $R^{-5/(8d_j-8)}$ the van der Corput estimate is not good enough to prove theorem 6.2 anymore. The reason is that the phase of the exponential sum in (6.6) is almost linear in n, as proposition 6.9 shows, and the oscillation is not captured by a second derivative test.

Throughout this section we will assume that R, M, U, U_1, U_2 and j are as in the statement of theorem 6.2, $UM \ge R^{5/8}$ (see comments in §6.5) and $M \le m < 2M$. Let $I_{m,\ell} = [U_1(m+\ell)^2, \min(U_2 m^2, N)]$, which we may assume non-empty by restricting the possible values of m, and define the quantities

$$\phi_\ell(n,m) = R\left(\frac{\partial h}{\partial n}(n, m+\ell) - \frac{\partial h}{\partial n}(n,m)\right),$$
$$\Phi_\ell(m) = \min_{x \in I_{m,\ell}} \|\phi_\ell(x,m)\|_{\mathbb{Z}}.$$

The function ϕ_ℓ is the derivative of the phase of the exponential sum in n appearing in T defined by (6.6). Since by proposition 6.7 this function is monotone in n, we can apply Kuzmin-Landau's lemma (proposition 4.3) yielding the bound

$$\left| \sum_{n \in I_{m,\ell}} e\big(R(h(n,m+\ell) - h(n,m))\big) \right| \ll \big(\Phi_\ell(m)\big)^{-1}.$$

Suppose we can find another bound H_ℓ for the same exponential sum, this time uniform in m, to apply in those cases when $\Phi_\ell \approx 0$. Then knowing very little about the distribution of the values $\Phi_\ell(m)$ we can find a good bound for T. The underlying idea here is to gain from some control of the spacing. In [**10**] and [**59**] this is accomplished via large sieve inequalities, while we introduce the spacing through the following simple result:

LEMMA 6.12. *Assume we have a finite sequence of points $a_m \in [0, \frac{1}{2}]$ satisfying for some $A, B \ge 0$ the following condition:*

$$\#\{m : a_m \le x\} \le A + Bx \qquad \text{for every} \quad 0 \le x \le 1/2.$$

Then for any $H > 0$ we have

$$\sum_m \min\{H, a_m^{-1}\} \le AH + B\big(1 + |\log H|\big).$$

Assuming, in our setting, that A_ℓ, B_ℓ and H_ℓ are functions of ℓ satisfying that $\ell A_\ell H_\ell$ and ℓB_ℓ are non-decreasing, and H_ℓ is bounded above and below by powers of R, it follows from this result that for any fixed $\epsilon > 0$,

$$T(U_1, U_2, M, L) \ll R^\epsilon (A_L H_L + B_L). \tag{6.9}$$

PROOF OF LEMMA 6.12. Let us say that the finite sequence is $0 \le a_1 \le a_2 \le \cdots \le a_N \le 1/2$, and assume $B > 0$ (as the case $B = 0$ is trivial). Note that, by hypothesis, $m \le A + Ba_m$. Let $f : [0, \frac{1}{2}] \to \mathbb{R}$ be any non-increasing function and extend it to the negative real numbers as the constant function $f(0)$. Then

$$\sum_m f(a_m) \le \sum_m f\left(\frac{m-A}{B}\right) \le B \int_{-A/B}^{1/2} f(x)\,dx = Af(0) + B\int_0^{1/2} f(x)\,dx.$$

The result follows applying this inequality with $f(x) = \min\{H, x^{-1}\}$. □

The upper bound H_ℓ will be either the trivial estimate UM^2, or the second term in the van der Corput estimate given by proposition 6.10 (the first one may be neglected in the range $U \ll R^{-5/(8d_j-8)}$, $UM \ge R^{3/8}$). The pair (A_ℓ, B_ℓ) will be given by one of the following two propositions.

PROPOSITION 6.13. *Assume $U^{d_j+1}M$ is small enough and $1 \le \ell \le UM$. Then*

$$\#\{m : \Phi_\ell(m) \le x\} \ll 1 + \frac{R\ell}{M^2} + M\left(1 + \frac{M^2}{R\ell}\right)x \qquad \text{for any} \quad 0 \le x \le 1/2.$$

PROOF. Choose for each pair (m, ℓ) a point $x_m \in I_{m,\ell}$ (depending implicitly on ℓ) satisfying

$$\Phi_\ell(m) = \|\phi_\ell(x_m, m)\|_{\mathbb{Z}}.$$

By the mean value theorem, $\phi_\ell(x_{m+1}, m+1) - \phi_\ell(x_m, m)$ equals

$$R\ell\frac{\partial^3 h}{\partial n \partial m^2}(x_1, y_1) + R\ell(x_{m+1} - x_m)\frac{\partial^3 h}{\partial n^2 \partial m}(x_2, y_2),$$

for some points (x_1, y_1), (x_2, y_2) lying in the rectangle

$$[U_1(m+\ell)^2, U_2(m+1)^2] \times [m, m+\ell+1].$$

The function x/y^2 over this rectangle satisfies

$$U_1(1 - 4M^{-1}) \le x/y^2 \le U_2(1 + 4M^{-1}),$$

and since $UM \ge R^{3/8}$ we have $|u_j - x_i/y_i^2| \asymp U$ for $i = 1, 2$. Using the estimates provided by propositions 6.7 and 6.8,

$$\phi_\ell(x_{m+1}, m+1) - \phi_\ell(x_m, m) \asymp \frac{R\ell}{M^3} + O\left(R\ell \cdot UM^2 \cdot \frac{U^{d_j}}{M^4}\right) \asymp \frac{R\ell}{M^3}, \tag{6.10}$$

the sign of the left hand side being always the same.

Since $M \le m < 2M$, we deduce from (6.10) that the number of integers k satisfying $|\phi_\ell(x_m, m) - k| \le 1/2$ for some m is at most a constant times $1 + R\ell M^{-2}$. On the other hand we also deduce from (6.10) that for each of those k and any $x \ge 0$

$$\#\{m : |\phi_\ell(x_m, m) - k| \le x\} \ll 1 + R^{-1}\ell^{-1}M^3 x.$$

Therefore,

$$\#\{m : \Phi_\ell(m) \le x\} \ll \left(1 + \frac{R\ell}{M^2}\right)\left(1 + \frac{M^3}{R\ell}x\right)$$

for every $0 \le x \le 1/2$. □

PROPOSITION 6.14. *Fix $\epsilon > 0$. For U small enough and $1 \le \ell \le UM$ we have*

$$\#\{m : \Phi_\ell(m) \le x\} \ll R^\epsilon(1 + R\ell U^{d_j+1} + M^2 x) \qquad (0 \le x \le 1/2).$$

PROOF. Let C_j the constant involved in proposition 6.9, and assume that we have

$$\left\| C_j \frac{R\ell}{m(m+\ell)} \right\|_{\mathbb{Z}} \leq x \qquad \text{for some} \quad x \geq 0.$$

This means that there exists an integer $k = k(m, \ell)$ satisfying

$$|C_j R\ell - km(m+\ell)| \leq m(m+\ell)x \ll M^2 x.$$

In particular, m must divide a certain integer $km(m+\ell)$ lying in the interval centered at $C_j R\ell$ of half-length a constant times $M^2 x$. Since there are $O(1+M^2x)$ of these integers, and each has at most $O(R^\epsilon)$ divisors, we conclude

$$\#\{m : \|C_j R\ell/(m(m+\ell))\|_{\mathbb{Z}} \leq x\} \ll R^\epsilon(1+M^2x).$$

Replacing x by $x + O\big(R\ell U^{d_j+1}M^{-2}\big)$ the result follows from proposition 6.9. □

This last argument is remarkably similar to the one used in §5.2 to prove Popov's result for the parabola, and later employed again in the same chapter for counting points inside elliptic paraboloids. They are, in some sense, the same argument. In chapter 5 it was used to show that the coefficient m/R of the quadratic form in n forming part of the phase function was very seldom close to rational numbers with big denominators. Recall we only did Poisson summation in one variable; if we had done it in every variable then this coefficient would essentially had been replaced by its inverse R/m (similarly to how theta functions transform, *cf.* §2.8). In this setting the same divisibility argument can be adapted to show that R/m is seldom close to a rational of small denominator. Now, here we were forced to apply a Weyl step to get rid of the function $r_2^*(n)$, gaining one derivative in the m variable, essentially replacing R/m by R/m^2 (nevermind the parameter ℓ). The same divisibility argument still shows that R/m^2 is seldom close to a rational number with small denominator, although for our purposes we only need to know that it is seldom close to an integer (k may be replaced by p/q in the previous proof to obtain a stronger result).

If we compare the spacing provided by propositions 6.13 and 6.14, we notice that the slope of the bound is much more step in the second case, but also in this case the independent term decreases to R^ϵ as $U \to 0$. This makes sense, the first proposition gains spacing from purely analytic methods, blind to whether the curve looks like a parabola or not. On the other hand, the second proposition is obtaining the spacing by using arithmetic properties of this curve, so it can only work well if the curve is really close to a parabola, and this happens precisely when we are really close to some u_j. Comparing the independent terms we might expect the bounds derived from proposition 6.14 to be sharper when $M^2U^{d_j+1} \ll R^{-\epsilon}$, and in particular when $U \ll R^{-5/(4d_j+4)-\epsilon}$. This is pretty close to the truth, as we see below in the statements of propositions 6.15 and 6.16, where the $+4$ in the denominator of the exponent is replaced by either -16 or $+24$. The proof of the first one uses exclusively the bounds derived from proposition 6.13, while the second one uses only the ones derived from proposition 6.14.

The following two propositions, together with proposition 6.11 in §6.5, complete the proof of theorem 6.2, and hence also the proof of theorem 6.1.

PROPOSITION 6.15. *If $U \ll R^{-5/(8d_j+8)}$ for a sufficiently small constant then theorem 6.2 holds when $d_j \leq 4$, or when $d_j \geq 5$ and $U \gg R^{-5/(4d_j-16)}$.*

PROOF. We apply proposition 6.4 to bound S with $L = R^{-3/4}U^2M^2$, which always lies in the interval $[1, UM]$. Using (6.9) with (A_L, B_L) given by proposition 6.13 (note the hypotheses imply $U^{d_j+1}M$ is small enough) we obtain

$$R^{-\epsilon}M^{-4}|S|^2 \ll R^{3/4} + \frac{UH_L}{M}\left(1 + \frac{RL}{M^2}\right) + U\left(1 + \frac{M^2}{RL}\right). \tag{6.11}$$

We choose either $H_L = UM^2$ or $H_L = R^{-1/8}U^{-(d_j+2)/2}M$ (second term in proposition 6.10) depending on whether $RL/M^2 \leq 1$ or not. In the first case, the right hand side of (6.11) may be bounded above by $R^{3/4} + U^2M + R^{-1/4}U^{-1}$, and using $M \leq R^{5/8}$ and $U \geq R^{-1/4}$ (from $UM \geq R^{3/8}$) we conclude $M^{-4}|S|^2 \ll R^{3/4+\epsilon}$. In the second case, the right hand side of (6.11) may be bounded above by $R^{3/4} + R^{1/8}U^{-(d_j-4)/2} + U$, which also leads to $M^{-4}|S|^2 \ll R^{3/4+\epsilon}$ under the hypotheses of this proposition. □

PROPOSITION 6.16. *Theorem 6.2 holds when* $U \ll R^{-5/(4d_j+24)}$.

PROOF. We proceed similarly as in the proof of proposition 6.15. We apply again proposition 6.4 to bound S with $L = R^{-3/4}U^2M^2$, and use (6.9) with (A_L, B_L) given by proposition 6.14 to obtain

$$R^{-\epsilon}M^{-4}|S|^2 \ll R^{3/4} + \frac{UH_L}{M}(1 + RLU^{d_j+1}) + UM. \tag{6.12}$$

We choose either $H_L = UM^2$ or $H_L = R^{-1/8}U^{-(d_j+2)/2}M$ depending on whether $RLU^{d_j+1} \leq 1$ or not. In the first case (6.12) shows that $M^{-4}|S|^2 \ll R^{3/4+\epsilon}$ is satisfied trivially, while in the second case the right hand side of (6.12) may be bounded above by $R^{3/4} + R^{1/8}U^{(d_j+6)/2}M^2 + UM$, which also leads to $M^{-4}|S|^2 \ll R^{3/4+\epsilon}$ under the hypothesis of this proposition. □

Appendix: toolbox

A.1. Poisson summation

Let $f : \mathbb{R} \to \mathbb{C}$ be a function with decay $f(x) = O(x^{-1-\epsilon})$ and uniformly ϵ-Hölder for some $\epsilon > 0$. Then

$$\sum_{n\in\mathbb{Z}} f(n) = \sum_{n\in\mathbb{Z}} \hat{f}(n).$$

To prove the formula above note that

$$\sum_{|n|\leq N} \hat{f}(n) = \int_{-\infty}^{+\infty} f(t)D_N(t)\,dt = \sum_{n\in\mathbb{Z}} \int_{-1/2}^{1/2} f(n-t)D_N(t)\,dt$$

where D_N is the Dirichlet kernel of order N. By Fubini we may interchange the last sum with the integral to obtain

$$\sum_{|n|\leq N} \hat{f}(n) = \int_{-1/2}^{1/2} g(-t)D_N(t)\,dt$$

where $g(x) = \sum_{n\in\mathbb{Z}} f(n+x)$. When we take the limit $N \to \infty$ the right hand side corresponds to the Fourier series of the periodic function g evaluated at 0, and hence converges to $g(0)$ provided that g has some regularity at this point. Since

$$|g(h) - g(0)| \leq \sum_{|n|\leq M} |f(n+h) - f(n)| + O(M^{-\epsilon}) \ll Mh^{\epsilon} + M^{-\epsilon},$$

taking $M = \lfloor h^{-\epsilon/(\epsilon+1)} \rfloor$ we see that g is $\epsilon^2/(\epsilon+1)$-Hölder at zero.

The d-dimensional version of the same formula states $\sum_{\vec{n}\in\mathbb{Z}^d} f(\vec{n}) = \sum_{\vec{n}\in\mathbb{Z}^d} \hat{f}(\vec{n})$ provided that $f(x) = O(x^{-d-\epsilon})$ and f is uniformly ϵ-Hölder in each variable for some $\epsilon > 0$. To prove this result one can either adapt the proof above or, maybe under slightly stronger regularity hypotheses to ensure $\hat{f}$ has enough decay, iterate the one-dimensional Poisson formula in each of the variables.

A.2. Summation by parts

The idea is simple: we know how to bound $\sum_{n=1}^{m} a_n$ for every m and we want to bound $\sum_{n=1}^{m} a_n b_n$, where the b_n vary "smoothly". We can do the following: let $S_N = \sum_{n=1}^{N} a_n$ and put $S_0 = 0$. Then $a_n = S_n - S_{n-1}$ and

$$\begin{aligned}\sum_{n=1}^{m} a_n b_n &= \sum_{n=1}^{m} (S_n - S_{n-1})b_n = \sum_{n=1}^{m} S_n b_n - \sum_{n=0}^{m-1} S_n b_{n+1} \\ &= \sum_{n=1}^{m-1} S_n(b_n - b_{n+1}) + S_m b_m.\end{aligned}$$

Now we can estimate the sum termwise, using a crude bound or the mean value theorem to estimate $|b_n - b_{n+1}|$ and the bounds we had for S_n. An equivalent

form, sometimes easier to estimate, for the sum on the right hand side when b_t is a differentiable function of t is $-\int_1^m S_{[t]}\,db_t$.

Exactly the same idea can be carried out in k variables. Let $S_{N_1,\dots,N_k} = \sum_{1\le n_i\le N_i} a_{n_1,\dots,n_k}$ with the convention $S_{N_1,\dots,N_k}=0$ if any $N_i\le 0$. Then to isolate $a_{n_1,\dots,n_k}$ we must apply a kind of inclusion-exclusion principle. For example, for $k=2$ we have $a_{n_1,n_2} = S_{n_1,n_2} - S_{n_1-1,n_2} - S_{n_1,n_2-1} + S_{n_1-1,n_2-1}$, as is easily shown by a diagram. In general,

$$a_{n_1,\dots,n_k} = \sum_{\delta_i\in\{0,1\}} (-1)^{\delta_1+\cdots+\delta_k} S_{n_1-\delta_1,\dots,n_k-\delta_k}.$$

Multiplying by $b_{n_1,\dots,b_k}$, summing again and performing in each sum the reindexing we obtain the general summation by parts formula:

$$\sum_{n_i\le m_i} a_{n_1,\dots,n_k} b_{n_1,\dots,n_k} = \sum_{n_i\le m_i-1} S_{n_1,\dots,n_k} \sum_{\delta_i\in\{0,1\}} (-1)^{\delta_1+\cdots+\delta_k} b_{n_1+\delta_1,\dots,n_k+\delta_k} + \Omega$$

where Ω are the boundary terms given by

$$\Omega = \sum_{\emptyset\neq\Pi\subset\{1,\dots,k\}} \sum_{\substack{n_i=m_i \text{ for } i\in\Pi\\ n_i\le m_i-1 \text{ for } i\notin\Pi}} S_{n_1,\dots,n_k} \sum_{\substack{\delta_i=0 \text{ for } i\in\Pi\\ \delta_i\in\{0,1\} \text{ for } i\notin\Pi}} (-1)^{\delta_1+\cdots+\delta_k} b_{n_1+\delta_1,\dots,n_k+\delta_k}.$$

Note the formula admits a more compact form, as this expression for $\Pi=\emptyset$ evaluates to the sum we have set apart above. Also, as with the unidimensional case, the rightmost sum can be turned into integrals for ease of estimation when b is a differentiable function of its subindices. In this case it equals:

$$(-1)^{k-\#\Pi}\left[\int_{n_i\le t_i\le n_i+1 \text{ for } i\notin\Pi}\cdots\int \left(\prod_{i\notin\Pi}\frac{\partial}{\partial t_i}\right) b_{t_1,\dots,t_k} \prod_{i\notin\Pi} dt_i\right]_{t_i=n_i \text{ for } i\in\Pi}.$$

A.3. Kernels of summability

Let a_n be a sequence of complex numbers summing a, *i.e.* $a=\sum_{n\ge0} a_n$. Suppose we have another sequence $b_n(t)$ depending on a parameter $t\in\mathbb{R}^+$, uniformly bounded in n and t, satisfying $\lim_{t\to\infty} b_n(t)=1$ for all $t>0$ and for some constant $C>0$,

$$\sum_{n\ge0} |b_{n+1}(t)-b_n(t)| < C \qquad \text{for any} \quad t>0.$$

Then

$$a = \lim_{t\to\infty} \sum_{n\ge0} a_n b_n(t).$$

Of course the point where we are taking the limit is unimportant. The usual Abel summation, for example, corresponds to $b_n(t)=t^n$ and $t\to 1^-$. A much more general theorem is provided by Zygmund in theorem III.1.2 of [**98**].

The proof of the result is as follows. Without loss of generality we may assume $a=0$, subtracting a constant from a_0 otherwise. Hence for every $\epsilon>0$ we may find some $N>0$ such that the partial sums $S_n=\sum_{m=0}^n a_m$ are bounded by ϵ for every $n\ge N$. Summing by parts,

$$\left|\sum_{n=1}^{\infty} a_n b_n(t)\right| \le \sum_{n=0}^{N-1} |S_n|\,|b_{n+1}(t)-b_n(t)| + \epsilon \sum_{n\ge N} |b_{n+1}(t)-b_n(t)|.$$

The second term is bounded by $C\epsilon$ while the first one goes to zero as $t\to\infty$.

A.4. Euler-Maclaurin formula

The Euler-Maclaurin formula is a powerful relation between sums and integrals which works in both ways: it can be used to estimate integrals by sums or, in our case, to estimate sums involving (hopefully) easier-to-estimate integrals. The formula states that for any $k \geq 0$, $a, b \in \mathbb{Z}$ and $f \in \mathcal{C}^{2k+1}([a,b])$,

$$\sum_{n=a}^{b} f(n) = \int_a^b f(t)\,dt + \frac{f(a)+f(b)}{2} + \sum_{m=1}^{k} \frac{B_{2m}}{(2m)!}\Big(f^{(2m-1)}(b) - f^{(2m-1)}(a)\Big) + \frac{1}{(2k+1)!}\int_a^b B_{2k+1}(\{t\})\, f^{(2k+1)}(t)\,dt.$$

The sum on the right hand side is understood to vanish for $k = 0$. The last term, although explicit, is usually regarded as the error term. The polynomial $B_n(x)$ is the n-th Bernoulli polynomial, defined inductively by $B_1(x) = x - 1/2$, $B_n'(x) = nB_{n-1}(x)$ and $\int_0^1 B_n(x)\,dx = 0$, and for $n \geq 2$, $B_n = B_n(0) = B_n(1)$ the n-th Bernoulli number. In particular, $B_{2k+1}(\{t\})$ is a 1-periodic function with vanishing integral on each period. If $f \in \mathcal{C}^{2k+2}([a,b])$, integrating by parts, the error term equals

$$\frac{B_{2k+2}}{(2k+2)!}\Big(f^{(2k+1)}(b) - f^{(2k+1)}(a)\Big) + O\Big(\operatorname*{var}_{a\leq x\leq b} f^{(2k+1)}(x)\Big)$$

where var stands for the total variation. The formula is also often employed with non-integer limits A and B, in which case it may be applied to $a = \lceil A \rceil$ and $b = \lfloor B \rfloor$.

To prove the formula, note it suffices to show

$$\frac{f(a)+f(a+1)}{2} = \int_a^{a+1} f(t)\,dt + \sum_{m=1}^{k} \frac{B_{2m}}{(2m)!}\Big(f^{(2m-1)}(a+1) - f^{(2m-1)}(a)\Big) + \frac{1}{(2k+1)!}\int_a^{a+1} B_{2k+1}(t-a)\, f^{(2k+1)}(t)\,dt,$$

as then summing this formula over $a, a+1, \ldots, b-1$ we obtain the formula above. The latter formula is just an exercise of integration by parts, starting from the integral $\int_a^{a+1} 1 \cdot f$, as $\frac{\partial}{\partial t} B_1(t-a) = 1$ and $\frac{\partial}{\partial t} n^{-1} B_n(t-a) = B_{n-1}(t-a)$. One has to use $B_n(0) = B_n(1) = B_n$ for $n \geq 2$, the n-th Bernoulli number, which vanishes for n odd. All these facts can be shown from the generating series $te^{xt}/(e^t - 1) = \sum_{n\geq 0} B_n(x)t^n/n!$.

Introducción y conclusiones[1, 2]

El objetivo inicialmente propuesto para esta tesis fue el de resolver varios problemas, pequeños pero con cierto interés, pertenecientes a la intersección entre la teoría analítica de números y el análisis armónico. Si tuvieramos sin embargo que elegir *a posteriori* un *leitmotiv* para esta exposición, sería sin duda la *función theta de Jacobi*:

$$\theta(z) = \sum_{n\in\mathbb{Z}} e^{\pi i n^2 z}. \tag{Ⅱ.1}$$

Esta función, claramente holomorfa en el semiplano superior, resulta ser una *forma modular*. Esto significa que satisface una ecuación funcional con respecto a la acción del grupo $\mathrm{SL}_2(\mathbb{Z})$ sobre el semiplano superior, y que es de crecimiento como mucho polinomial cuando $\Im z \to 0^+$.

Jacobi fue el primero en estudiar sistemáticamente las propiedades de esta función, a raíz de su trabajo [**63**] sobre *integrales elípticas*. Una integral elíptica es una función de la forma

$$\int_c^x R\Big(t, \sqrt{P(t)}\Big)\, dt \tag{Ⅱ.2}$$

donde c es una constante, R una función racional y P un polinomio de grado tres o cuatro. Estas integrales aparecen de manera natural al intentar calcular la longitud de un arco de elipse (de aquí la nomenclatura), así como en ciertos problemas de índole física, incluyendo la evolución de la distancia al Sol de un planeta y la del ángulo de un péndulo, en función del tiempo.

Toda integral elíptica (Ⅱ.2) admite una expresión cerrada en términos de funciones elementales si a estas añadimos tres familias de funciones especiales: *las integrales elípticas incompletas de primera, segunda y tercera especie*. En particular, las de primera especie, son funciones de la forma

$$F(x;k) = \int_0^x \frac{dt}{\sqrt{(1-t^2)(1-k^2t^2)}},$$

donde el parámetro k recibe el nombre de *módulo*.

Resulta que en lugar de estudiar directamente la función $F(x;k)$ es mucho más conveniente centrarse en su función inversa. Esto es análogo a lo que pasa con las funciones logaritmo o arcoseno, que ambas admiten definiciones sencillas en términos de una integral, pero sus inversas, la exponencial y el seno, disfrutan de mejores propiedades analíticas. En particular, son funciones univaluadas y enteras en todo

[1] Este capítulo se incluye para cumplir con la normativa de la Universidad Autónoma de Madrid referente a tesis presentadas en un idioma extranjero. Sintetiza el contenido del capítulo introductorio y de las secciones §3.2, §5.1 y §6.1

[2] This chapter is included to comply with the regulations of the Universidad Autónoma de Madrid regarding dissertations written in a foreign language. It synthesizes the contents of the introductory chapter and of sections §3.2, §5.1 and §6.1.

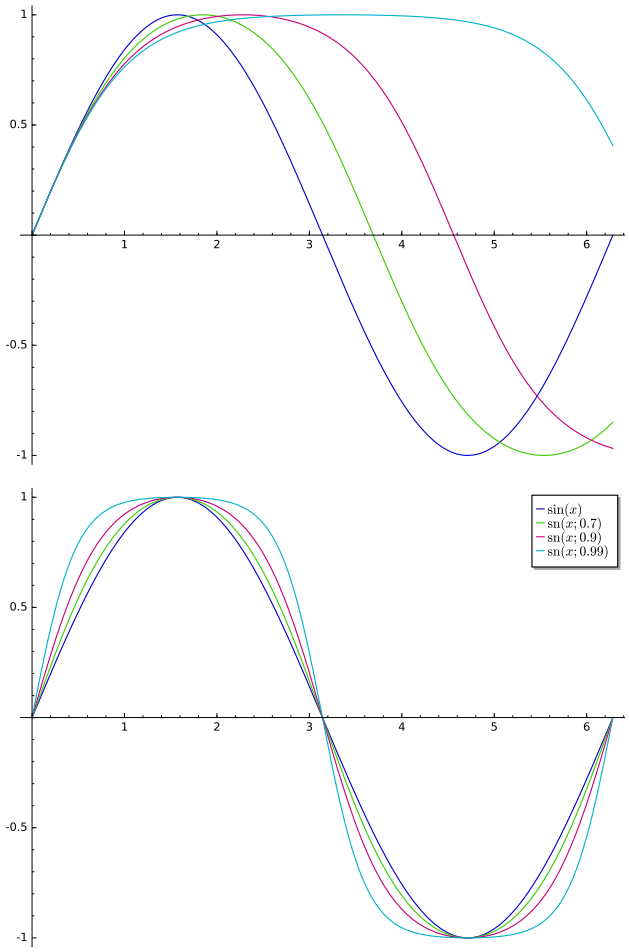


FIGURA **II**.1. La función seno elíptico para varios valores del módulo k. En la imagen de abajo la variable x ha sido reescalada para cada valor de k con el fin de que todos los periodos tengan longitud 2π.

el plano complejo. De manera análoga, el *seno elíptico* sn, definido por la relación $F(\mathrm{sn}(x,k);k) = x$, es una función univaluada meromorfa en todo el plano complejo. Esta función fue estudiada por Legendre y Abel, y más tarde en profundidad por Jacobi. En la figura **II**.1 se puede apreciar su gráfica para x real: el seno elíptico es un función periódica cuyo periodo depende del valor del módulo k.

Sorprendentemente el seno elíptico, para $k \neq 0$ (ya que para $k = 0$ coincide con el seno usual), tiene un segundo periodo complejo. Es a raíz de esto que a las funciones meromorfas en el plano complejo que tienen dos periodos linealmente independientes sobre $\mathbb{R}$ se las llama *funciones elípticas*. Estas funciones son asombrosamente rígidas, y de hecho las únicas funciones elípticas enteras son las constantes. Se deduce de este hecho que siempre que tengamos dos funciones elípticas cuyos periodos coincidan y cuyos ceros y polos también, una de ellas ha de ser un múltiplo constante de la otra. Esto constituye una poderosa herramienta para probar identidades que *a priori* no son en absoluto obvias.

Jacobi se dió cuenta de que la función de dos variables

$$\Theta(z;\tau) = \sum_{n\in\mathbb{Z}} q^{n^2} e^{2\pi i n z} \qquad \text{donde} \quad q = e^{i\pi\tau}$$

para τ fijo en el semiplano superior es 1-periódica en la variable z, y casi τ-periódica en esta misma variable. Más concretamente cumple

$$\Theta(z+1;\tau) = \Theta(z;\tau) \qquad \text{y} \qquad \Theta(z+\tau;\tau) = q^{-1}e^{-2\pi iz}\Theta(z;\tau),$$

y de aquí se deduce que el cociente $\Theta(z+\tau/2;\tau)/\Theta(z+(\tau+1)/2;\tau)$ es una función elíptica de periodos 1 y 2τ. Tras ajustar constantes, el valor de τ y dilatar adecuadamente la variable z, Jacobi prueba usando la rigidez de las funciones elípticas que este cociente provee una expresión alternativa para el seno elíptico. Esta nueva expresión resulta útil tanto a la hora de probar formalmente ciertas propiedades de sn como a la hora de calcular valores numéricamente, ya que Θ viene dada por una serie de convergencia exponencial.

Jacobi además se percató de que la funcion Θ, siendo en la variable $\tau/2$ una serie de Fourier soportada en los cuadrados, se puede emplear para obtener información sobre este conjunto de enteros. Fue mediante esta conexión que fue capaz de probar su famoso teorema de los cuatro cuadrados:

Teorema (Jacobi). *El número de formas de representar un entero n como suma de cuatro cuadrados coincide con ocho veces la suma de sus divisores si n es impar, y veinticuatro veces la suma de sus divisores impares si n es par.*

Este teorema se puede reformular como una identidad entre dos series generatrices en q, una de las cuales viene dada por $\left(\Theta(0;\tau)\right)^4$. Probar que son la misma función, sin embargo, no es sencillo ya que ninguna de las dos funciones depende de la variable z, en la cual uno podría explotar la rigidez de las funciones elípticas. Jacobi consiguió solventar esto a base de pasar por otras identidades intermedias que involucran funciones que sí son elipticas, y luego especializando $z = 0$ [**63**]. La manera moderna de probar el mismo resultado se basa directamente en la ley de transformación en la variable τ, que en este caso coincide con la de la forma modular $\theta(\tau) = \Theta(0;\tau)$ introducida en (**II**.1) (*cf.* §7.4 of [**82**]). De hecho, el adjetivo modular viene de aquí: de la relación con k, el parámetro del seno elíptico, ya que cuando se escribe sn como cociente de funciones theta el módulo k y la variable τ quedan relacionados precisamente por una función que cumple esta ley de transformación.

Además de con las integrales elípticas, las formas modulares también están íntimamente relacionadas con las *curvas elípticas*. De hecho, fueron las integrales elípticas las que dieron origen a estas últimas. Tal y como el seno y el coseno parametrizan el círculo, y cumplen fórmulas de adición que se pueden usar para dotar al círculo de su estructura de grupo usual; de la misma manera el seno elíptico, junto con dos funciones trigonométricas elípticas más, parametrizan una curva en un espacio tridimensional, y cumplen leyes de adición que dotan a dicha curva de una estructura de grupo. Estas curvas luego se vio que eran equivalentes a las curvas planas dadas por ecuaciones de la forma $y^2 = x^3 + ax + b$ con $4a^3 + 27b^2 \neq 0$, forma más conveniente. Si se añaden los puntos complejos que "faltan" estas curvas resultan equivalentes a toros obtenidos al cocientar el plano complejo por un retículo, y aquí de nuevo hacen aparición las funciones elípticas como aquellas funciones que cocientan bien y "viven" en la curva elíptica. Al final el adjetivo "elíptico" deja patente la interrelación entre estos objetos, aunque a menudo se olvide mencionar el origen común que tienen en el cómputo de la longitud de ciertos arcos de elipses.

La noción de forma modular, hoy en día, engloba una rica familia de funciones que aparecen por doquier en teoría de números. Concretemos más su definición: una forma modular es una función holomorfa $f : \mathbb{H} \to \mathbb{C}$, de crecimiento a lo sumo

polinomial cuando $\Im z \to 0^+$, que se transforma de la siguiente manera:

$$(\mathbf{II}.3) \qquad f(\gamma z) = \mu_\gamma (cz+d)^r f(z) \qquad \text{para todo} \quad \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma,$$

donde Γ es un subgrupo de índice finito de $\mathrm{SL}_2(\mathbb{Z})$, $\gamma z = (az+b)/(cz+d)$, μ_γ es una constante unimodular dependiendo de γ y al real positivo r (unívocamente determinado por f) se lo denomina el *peso* de la forma modular. De esta definición se deduce (ver capítulo 2) que f posee un desarrollo como serie de Fourier

$$(\mathbf{II}.4) \qquad f(z) = \sum_{n+\kappa_\infty \geq 0} a_n e^{2\pi i (n+\kappa_\infty) z/m_\infty}$$

donde n recorre los enteros, $0 \leq \kappa_\infty < 1$ y m_∞ es un entero positivo. Los coeficientes $a_n \in \mathbb{C}$ son de crecimiento a lo sumo polinomial, y por tanto después de integrar formalmente suficientes veces esta serie converge uniformemente cuando $z \in \mathbb{R}$ a una función continua en la recta real. Generalizando esto, definamos para $\alpha > 0$ la serie formal

$$(\mathbf{II}.5) \qquad f_\alpha(z) = \sum_{n+\kappa_\infty > 0} \frac{a_n}{(n+\kappa_\infty)^\alpha} e^{2\pi i (n+\kappa_\infty) z/m_\infty}.$$

La serie f_α, de converger, es esencialmente una "integral α-ésima" de la función f.

La acción del subgrupo Γ sobre $\mathbb{H}$ que aparece en la definición de forma modular se extiende trivialmente a una acción sobre $\mathbb{R} \cup \{\infty\}$, y esto le otorga a f_α un aspecto fractal muy particular. Por ejemplo, en la figura **II**.2 hemos incluido el grafo de la función

$$\varphi(x) = \sum_{n \geq 1} \frac{\sin(n^2 \pi x)}{n^2},$$

que con la notación de arriba coincide con $\frac{1}{2}\Im\theta_1$. Esta función tiene una larga y controvertida historia, siendo mencionada por primera vez por Weierstrass en una charla que da frente a la academia de ciencias de Berlín en 1872. En esta charla, centrada en funciones con poca regularidad, Weierstrass comenta que φ fue propuesta por Riemann a sus estudiantes como ejemplo de una función no diferenciable en ningún punto, pero que, sin embargo, él no ve sencillo probarlo y prefiere presentar de manera alternativa el (ahora más conocido) ejemplo

$$\sum_{n \geq 0} a^n \cos(b^n \pi x)$$

para a, b satisfaciendo $0 < a < 1$, b un entero positivo impar y $ab > 1 + 3\pi/2$.

A raíz de este comentario muchos autores se han interesado por la regularidad de φ, y por la veracidad de las afirmaciones de Weierstrass. En particular, Butzer y Stark [**13**] analizan el tema a partir de unas cartas que fueron encontradas de Christoffel dirigida a Prym (el primero antiguo estudiante de Riemann), en las que comentan el asunto, y la evidencia apunta a que la función φ jamás fue mencionada por Riemann, y Weierstrass debió hacerse la idea equivocada a raíz de alguna confusión con alguno de los estudiantes de Riemann. A pesar de esto, como los propios autores de este artículo lo expresan, la evidencia no es sólida y quién si no Riemann podía tener el ingenio necesario para concebir un ejemplo así.

En cualquier caso la función φ ha pasado a la historia conocida como "el ejemplo de Riemann de una función no diferenciable" (abreviadamente "el ejemplo de Riemann"). El primero en publicar algún resultado sobre la misma fue Hardy, quien prueba en 1916 [**42**], casi cincuenta años después de la intervención de Weierstrass,

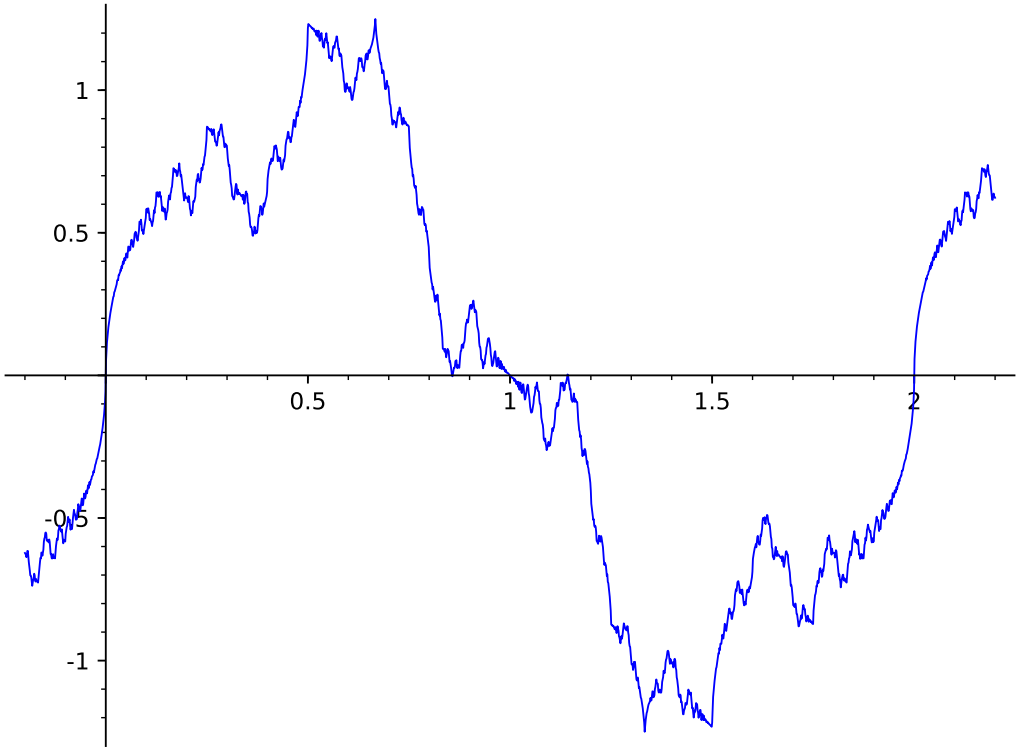


FIGURA **II**.2. El aspecto del "ejemplo de Riemann" φ.

que φ no tiene derivada en ningún irracional, ni tampoco en ningún racional salvo, tal vez, en aquellos de la forma impar/impar ó par/$(4n+3)$. Otros cincuenta años tendrían que pasar para que Gerver completara el resultado de Hardy [**35, 36**] mostrando que φ no es diferenciable en aquellos racionales de la forma par/$(4n+3)$ pero sí lo es en los de la forma impar/impar, teniendo en estos derivada $-\pi/2$. Esto último se puede apreciar en la figura **II**.2 aunque, por supuesto, en la época de Weierstrass o de Hardy era imposible conseguir una gráfica tan detallada de φ.

El resultado de Hardy se basa en una ingeniosa transformada integral que, aplicada a φ, devuelve θ. Esta es esencialmente un inverso de la integral de Riemann-Liouville. Utilizando esta transformada como nexo, Hardy relaciona la regularidad de φ en un punto de la recta real con el comportamiento de θ en el semiplano superior cerca de este punto. Esto, más la ecuación funcional (**II**.3) que liga el tamaño de θ cerca de un punto real con propiedades diofánticas del real en cuestión (objetivo del artículo [**45**] desarrollado con anterioridad por Hardy en compañía de Littlewood) le permiten a Hardy derivar su teorema. Esta misma idea ha sido rescatada recientemente bajo el formalismo de la *transformada ondícula*, permitiendo a Holschneider, Tchamitchian y Jaffard refinar los resultados de Hardy y Gerver [**52, 64, 65**] dando información más precisa sobre qué condiciones Hölder cumple la función φ en cada punto. En concreto, determinan el llamado *exponente Hölder puntual*

$$\beta(x_0) = \sup\{s \,:\, f \in \mathcal{C}^s(x_0)\}$$

donde $\mathcal{C}^s(x_0)$ denota el espacio de aquellas funciones continuas cumpliendo para algún polinomio P la desigualdad

$$|f(x) - P(x-x_0)| \ll |x-x_0|^s \qquad \text{cuando} \quad x \to x_0.$$

Jaffard va más allá, y es capaz de determinar también el llamado *espectro de singularidades* de φ [**64**]. Este, para una función continua, se define como la aplicación $d : [0,\infty) \to [0,1]\cup\{-\infty\}$ que asocia a cada $\delta > 0$ la dimensión de Hausdorff del conjunto $\{x : \beta(x) = \delta\}$ si este conjunto es no vacío y $-\infty$ en caso contrario. Jaffard se da cuenta de que β en el caso del "ejemplo de Riemann" en los puntos irracionales depende de cómo de bien se pueden aproximar dichos puntos por racionales de la forma impar/impar, y es capaz de adaptar a tal efecto el clásico resultado de Jarník y Besicovitch [**66**] para determinar la dimensión de estos conjuntos.

Por otro lado, Duistermaat en [**24**] encuentra un enfoque alternativo para tratar la regularidad de φ, especialmente cerca de los racionales. Esto lo consigue integrando la ecuación funcional (**II**.3) para obtener una *ecuación funcional aproximada*, válida para φ, con un término de error que es posible controlar cerca de ciertos racionales. De aquí deduce que alrededor de algunos racionales, a un lado o a los dos, aparecen singularidades de tipo raíz cuadrada, las cuales son apreciables a simple vista en la figura **II**.2 (por ejemplo, alrededor de 0 y a la izquierda de 1/2). Más aún, la ecuación funcional explica la autosemejanza del grafo cerca de algunos racionales (0, por ejemplo), alrededor de los cuales aparece una versión deformada del propio grafo de φ repitiéndose con amplitud decreciente.

Ambos enfoques (tanto el de Hardy como el de Duistermaat) tienen en común que el ingrediente principal es la ecuación funcional que cumple θ por ser una forma modular (**II**.3). Cabe la pregunta de si para otras formas modulares se puede hacer algo parecido. La respuesta es afirmativa, y estas técnicas con las adecuadas modificaciones se pueden aplicar para estudiar en general la función f_α definida por (**II**.5). Esta investigación fue comenzada por F. Chamizo en [**14**], y continuada por Chamizo, Petrykiewicz y Ruiz-Cabello en [**19**] y por Ruiz-Cabello en [**83**], trabajos en los que se consiguió determinar el exponente Hölder puntual bajo restricciones muy fuertes en el tipo de formas consideradas y en los valores de α y del peso de la forma modular r. Esto se debe por un lado a que emplearon la misma definición de *ondícula* que Jaffard, cuando una versión ligeramente modificada resulta más adecuada para tratar este problema, y por otro a que sólo consideraron la ecuación funcional aproximada en una versión muy rudimentaria. El autor consiguió en [**80**], con la inestimable ayuda de F. Chamizo, subsanar estos déficits y obtener los teoremas que detallamos a continuación.

Nos hace falta introducir un poco de notación. Dada una matriz $\gamma \in \mathrm{GL}^+(\mathbb{R})$ definimos la función

$$f^\gamma(z) = (\det\gamma)^{r/2}\frac{f(\gamma z)}{\big(j_\gamma(z)\big)^r}$$

donde $j_\gamma(z)$ denota el denominador de la transformación fraccional lineal asociada a γ. Si el grupo $\gamma^{-1}\Gamma\gamma \cap \mathrm{SL}_2(\mathbb{Z})$ vuelve a ser un subgrupo de índice finito, se deduce de la definición de forma modular (ver capítulo 2) que la función f^γ es, de nuevo, una forma modular para este nuevo grupo. En particular admite un desarrollo de Fourier (**II**.4) que tiene asociada una integral formal (**II**.5). A esta última la denotamos por f_α^γ. Si la serie de Fourier (**II**.4) asociada a f^γ carece de término independiente se dice que f es *cuspidal en* $\gamma\infty$, y si es cuspidal en todo racional se dice que f es una *forma cuspidal*. Aquí, aunque no es estándar, también diremos por conveniencia que el racional $\gamma\infty$ es (o no) cuspidal para f. Además establecemos $\alpha_0 = r/2$ si f es una forma cuspidal y $\alpha_0 = r$ en caso contrario.

Teorema (Regularidad global). *Sea $\alpha > 0$. Se cumple:*

(*i*) *Si* $\alpha \leq \alpha_0$ *la serie formal* (**II**.5) *definiendo* f_α *diverge en un conjunto denso.*

(*ii*) *Si* $\alpha > \alpha_0$ *la serie formal* (**II**.5) *definiendo* f_α *converge uniformemente a una función continua en toda la recta real. Además,* f_α *admite* $\lceil \alpha - \alpha_0 \rceil - 1$ *derivadas, y la última derivada es* $\{\alpha - \alpha_0\}$*-Hölder continua si* $\alpha - \alpha_0 \notin \mathbb{Z}$ *y* s*-Hölder continua para todo* $s < 1$ *en caso contrario.*

(*iii*) *Si* $0 < \alpha - \alpha_0 \leq 1$ *entonces ni* f_α*, ni su parte real ni imaginaria, son derivables con continuidad en ningún intervalo* I*.*

Teorema (Regularidad local en los racionales). *Sea* $\alpha > \alpha_0$ *y* x *un número racional, y sea* $\beta(x)$ *el exponente Hölder puntual de* f_α*,* $\Re f_\alpha$ *ó* $\Im f_\alpha$*. Entonces* $\beta(x) = 2\alpha - r$ *si* f *es una forma cuspidal y* $\beta(x) = \alpha - r$ *en caso contrario. Si además* $0 < \alpha - \alpha_0 \leq 1$ *entonces* f_α *(resp.* $\Re f_\alpha$*,* $\Im f_\alpha$*) no es diferenciable en ningún racional que no sea cuspidal para* f*. Si* x *es cuspidal para* f *entonces* f_α *es diferenciable en* x *si y sólo si* $\alpha > (r+1)/2$*, y en este caso la derivada viene dada por*

$$f'_\alpha(x) = \frac{(2\pi)^\alpha}{(im)^\alpha \Gamma(\alpha)} \int_{(x)} (z-x)^{\alpha-1} f'(z)\, dz,$$

donde (x) *denota la semirrecta vertical que conecta* x *con* $i\infty$*.*

La regularidad en los irracionales depende de cómo de bien se aproximan estos por racionales que no sean cuspidales para f. Más concretamente, de la siguiente cantidad:

$$\tau_x := \sup\left\{\tau \,:\, \left|x - \frac{p}{q}\right| \ll \frac{1}{q^\tau} \text{ para infinitos racionales } \frac{p}{q} \text{ no cuspidales}\right\}.$$

Siempre se tiene la desigualdad $\tau_x \geq 2$ (ver prop. 2.3) y si $\tau_x = \infty$ establecemos la convención $1/\tau_x = 0$. Bajo estas consideraciones,

Teorema (Regularidad local en los irracionales). *Sea* $\alpha > \alpha_0$ *y* x *un número irracional, y sea* $\beta(x)$ *el exponente Hölder puntual de* f_α*,* $\Re f_\alpha$ *ó* $\Im f_\alpha$*. Si* f *es una forma cuspidal entonces* $\beta(x) = \alpha - r/2$*. En caso contrario,*

$$\beta(x) = \alpha - \left(1 - \frac{1}{\tau_x}\right) r.$$

Además de estos teoremas de regularidad también somos capaces de probar una ecuación funcional aproximada, al estimo de la de Duistermaat, que permite extraer información precisa sobre el comportamiento de las integrales fraccionarias f_α cerca de los números racionales.

Teorema (Ecuación funcional aproximada). *Sea* $\sigma \in \mathrm{SL}_2(\mathbb{R})$ *una matriz satisfaciendo que* f^σ *es una forma modular y que* $x_0 = \sigma\infty \in \mathbb{Q}$*. Asumamos además que el elemento inferior izquierdo de* σ *es negativo. Entonces existen dos constantes reales no nulas* A*,* B *con* $B > 0$*, dependiendo de* σ*, satisfaciendo:*

$$f_\alpha(x) = Ai^{-\alpha} f(x_0)\phi(x - x_0) + B|x - x_0|^{2\alpha}(x - x_0)^{-r} f_\alpha^\sigma(\sigma^{-1}x) + E(x)$$

donde $f(x_0) = \lim_{\Im z \to \infty} f^\sigma(z)$ *y*

$$\phi(x) = \begin{cases} x^{\alpha-r} & \text{si } \alpha - r \notin \mathbb{Z}, \\ x^{\alpha-r}\log x & \text{si } \alpha - r \in \mathbb{Z}. \end{cases}$$

El término de error $E(x)$ *es diferenciable con continuidad en* $\mathbb{R} \setminus \{x_0\}$ *y pertenece al espacio* $\mathcal{C}^{2\alpha-r+1}(x_0)$*.*

Como se ha mencionado arriba, también podemos generalizar el resultado de Jaffard determinando el espectro de singularidades d para f_α en general. Cuando la imagen de d no es discreta se dice que la función en cuestión es *multifractal*.

Teorema (Espectro de singularidades). *Sea d el espectro de singularidades de f_α, $\Re f_\alpha$ o de $\Im f_\alpha$. Entonces:*

(*i*) *Si f es una forma cuspidal:*

$$d(\delta) = \begin{cases} 1 & \text{si } \delta = \alpha - r/2, \\ 0 & \text{si } \delta = 2\alpha - r, \\ -\infty & \text{en caso contrario.} \end{cases}$$

(*ii*) *Si f no es una forma cuspidal:*

$$d(\delta) = \begin{cases} 2 + 2\frac{\delta-\alpha}{r} & \text{si } \alpha - r \leq \delta \leq \alpha - r/2, \\ 0 & \text{si } \delta = 2\alpha - r \text{ y } f \text{ es cuspidal en algún racional,} \\ -\infty & \text{en caso contrario.} \end{cases}$$

Las funciones f_α, $\Re f_\alpha$ y $\Im f_\alpha$ son, por tanto, multifractales si y sólo si f no es cuspidal.

Todos estos teoremas fueron publicados en el artículo "On the regularity of fractional integrals of modular forms" y las pruebas aparecen detalladas en el capítulo 3 de esta tesis. De hecho, en dicho capitulo no solo determinamos el exponente Hölder puntual, sino que además complementamos estos resultados determinando dos exponentes más relacionados, que miden diferentes aspectos locales de la regularidad de f_α. Estos exponentes aparecen en las investigaciones previas realizadas por Chamizo, Petrykiewicz y Ruiz-Cabello [**19**].

El resto de la tesis versa sobre problemas de conteo de puntos del retículo. El objetivo de los mismos, en una formulación bastante general, es la de estimar el número de puntos con coordenadas enteras que quedan dentro de una región de $\mathbb{R}^d$ que depende de uno o más parámetros, según estos parámetros varían. Nosotros nos vamos a centrar en una familia particular de estos problemas: estamos interesados en contar puntos de coordenadas enteras en RK, donde $K \subset \mathbb{R}^d$ es una región convexa fija que queda dilatada por el factor $R \to \infty$. Denotemos por $\mathcal{N}(R)$ el número de puntos de coordenadas enteras que queda dentro de RK para cada $R > 1$. En particular nos interesa particularmente el exponente

$$\alpha_K = \text{ínf}\,\{\alpha > 0 : \mathcal{N}(R) - |K|R^d = O(R^\alpha)\},$$

donde $|K|$ denota la medida de Lebesgue d-dimensional de K.

El origen de estos problemas se ubica en la prueba de la *fórmula del número de clases de Dirichlet*. Dicha fórmula, publicada por Dirichlet en 1839, en el caso de discriminante $d < 0$ corresponde a la identidad

$$h(d) = \frac{w}{2\pi}|d|^{1/2}L(1, \chi_d) \tag{II.6}$$

donde $h(d)$ es el número de clases asociado al cuerpo de números $\mathbb{Q}(\sqrt{d})$, el carácter χ_d viene dado por el símbolo de Kronecker $\left(\frac{d}{\cdot}\right)$ y w vale 6 para $d = -3$, vale 4 para $d = -4$ y vale 2 en el resto de los casos. Por supuesto en aquella época la teoría de cuerpos de números estaba aún mayormente por desarrollar, pero el número de clases se entendía como el número de formas cuadráticas $ax^2 + bxy + cy^2$ con coeficientes

enteros y discriminante $b^2 - 4ac = d$ que existen módulo relación de equivalencia por matrices en $\mathrm{SL}_2(\mathbb{Z})$. Esta misma definición es la que lleva más o menos directamente a una prueba de la identidad (**II**.6), esencialmente aplicando el método de la hipérbola de Dirichlet a la suma de carácteres $w \sum_{m|n} \left(\frac{d}{m}\right)$, que proporciona el número $R(n)$ de representaciones del entero n por un conjunto de representantes completo de las clases de equivalencia de formas cuadráticas de discriminante d. El argumento completo puede ser consultado en la versión en inglés de la introducción de esta tesis o en el capítulo §6 del libro de Davenport [**23**]. El punto clave está en interpretar geométricamente la cantidad $\sum_{n \leq N} R(n)$ como el número de puntos de coordenadas enteras dentro de las $h(d)$ elipses dadas por $Q_i(x, y) \leq N$, donde Q_i recorre dichos representantes, cantidad que asintóticamente crece como la suma de las áreas delimitadas por dichas elipses.

La fórmula del número de clases (**II**.6), al menos para el caso de discriminante negativo, era conocida por Gauss con anterioridad. De hecho, el lector puede comprobar que esta aparece en el artículo [**34**], publicado dos años antes que el trabajo de Dirichlet. De hecho, se piensa que Gauss estaba al tanto de dicha fórmula desde hacía muchos años, pero su lema "pauca sed matura" (pocos, pero maduros) le impedía publicar los resultados hasta haber extraído el máximo partido de los mismos. En particular, en este artículo, para probar la fórmula Gauss da un argumento elemental demostrando que cuando K es una elipse el exponente α_K definido anteriormente está acotado superiormente por 1. Este argumento consiste en cortar el plano en cuadrados de lado uno centrados en los puntos de coordenadas enteras; estableciendo una relación biunívoca entre cada cuadrado de área unidad y su punto central. Al final, contar puntos de coordenadas enteras contenidos en la elipse es casi como contar cuadrados de lado uno contenidos en la elipse, excepto por aquellos cuadrados que tocan el borde de la elipse. La cantidad de estos cuadrados "malos" es del mismo orden que el diámetro que la elipse, y por tanto crece como R, en contraposición al área que crece como R^2. Este mismo argumento aplicado en general a un cuerpo convexo K cualquiera d-dimensional con frontera suave muestra $\alpha_K \leq d - 1$.

En honor a Gauss, el problema de determinar α_K cuando K es el círculo unidad del plano centrado en el origen recibe el nombre de *problema del círculo de Gauss*. Este problema no solo ha atraído una gran atención, sino que sigue abierto en la actualidad. El primero en mejorar el resultado de Gauss fue Sierpiński [**89**], quien en 1906 usando ideas de Voronoï prueba $\alpha_K \leq 2/3$. La cota para α_K se ha ido mejorando lentamente; actualmente la mejor conocida es $\alpha_K \leq 517/824$ obtenida por Bourgain y Watt en 2017 [**11**]. Por otro lado, Hardy y Landau en 1915 [**41, 73**] prueban independientemente $\alpha_K \geq 1/2$, estableciendo $\alpha_K = 1/2$ como la conjetura más extendida hasta la actualidad.

Hoy en día existen multitud de artículos en la literatura en los que se obtienen cotas más o menos fuertes para α_K cuando K pertece a diversas familias concretas de cuerpos convexos. La mayor parte de estos resultados hacen uso de la transformada de Fourier como primer paso, en la forma de sumación de Poisson, para transformar el problema de acotar el término de error $\mathcal{N}(R) - |K|R^d$ por el de acotar una *suma exponencial*. Con el fin de ilustrar estas ideas esbozamos a continuación una prueba moderna del anteriormente mencionado resultado de Sierpiński para el círculo. Notemos que si χ_R es la función característica del círculo de radio R centrado en el origen, la suma $\sum_{\vec{n}} \chi_R(\vec{n})$ coincide justamente con $\mathcal{N}(R)$, mientras que $\hat{\chi}_R(\vec{0}) = |K|R^d$, con

lo que podemos pensar que el término de error viene dado por $\sum_{\vec{n}\neq\vec{0}} \hat{\chi}_R(\vec{n})$. Sin embargo la poca regularidad de la función χ_R impide que esta última suma converja, haciendo falaz la aplicación de la fórmula de sumación de Poisson. La solución es regularizar antes χ_R convolucionando con una función suave de soporte compacto. En efecto, elijamos para cierto $h = h(R) \leq 1$ una función meseta radial $\eta \in \mathcal{C}^\infty(\mathbb{R}^2)$ satisfaciendo

$$\eta \geq 0, \qquad \int \eta = 1 \qquad \text{y} \qquad \operatorname{supp} \eta \subset B(0,h).$$

Para cualquier $\epsilon > 0$, la suma $\sum_{\vec{n}} \chi_R * \eta(\vec{n})$ se ha modificado en a lo sumo $O(hR^{1+\epsilon})$, ya que la diferencia con $\sum_{\vec{n}} \chi_R(\vec{n})$ queda mayorada por el número de puntos de coordenadas enteras en la corona circular de radios $R+h$ y $R-h$; es decir, mayorada por

$$\sum_{(R-h)^2 \leq m \leq (R+h)^2} r_2(m)$$

donde $r_2(m)$ denota el número de maneras de escribir m como suma de dos cuadrados. La función r_2 cumple la cota $r_2(m) \ll m^\epsilon$ para todo $\epsilon > 0$ (véase §16.9 de [**46**]), con lo que queda justificada la afirmación anterior. Por tanto,

$$\begin{aligned}\mathcal{N}(R) + O(hR^{1+\epsilon}) &= \sum_{\vec{n}\in\mathbb{Z}^2} \chi_R * \eta(\vec{n}) = \pi R^2 + \sum_{\vec{0}\neq\vec{n}\in\mathbb{Z}^2} \hat{\chi}_R(\vec{n}) \cdot \hat{\eta}(\vec{n}) \\ &= \pi R^2 + R \sum_{n\geq 1} r_2(n) \hat{\eta}\big(\sqrt{n}\big) \frac{J_1\big(2\pi R\sqrt{n}\big)}{\sqrt{n}},\end{aligned}$$

donde hemos escrito $\hat{\eta}\big(\sqrt{n}\big)$ en lugar de $\hat{\eta}\big(\sqrt{n}, 0\big)$ y J_1 denota la función de Bessel de primera especie. Sustituyendo la archiconocida estimación asintótica (cap. VII de [**94**])

$$J_1(x) \sim \sqrt{\frac{2}{\pi x}} \cos\left(x - \frac{\pi}{4}\right) \ll \frac{1}{\sqrt{x}}, \tag{**II**.7}$$

obtenemos

$$\begin{aligned}\mathcal{N}(R) + O(hR^{1+\epsilon}) &= \pi R^2 + O\left(R^{1/2} h^{-5\epsilon/2} \sum_{1\leq n\leq h^{-2-2\epsilon}} \frac{1}{n^{3/4}}\right) + O(h^{-\epsilon} R^\epsilon) \\ &= \pi R^2 + O\Big(h^{-\frac{1}{2}-3\epsilon} R^{1/2}\Big).\end{aligned}$$

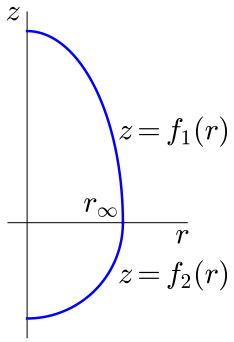
Basta ahora elegir $h = R^{-1/3}$.

El mismo argumento permite obtener cotas mejores si se aprovecha la cancelación proveniente del signo del coseno en (**II**.7). En general, cuando $K \subset \mathbb{R}^d$ con $d \geq 2$ se puede proceder de la misma manera siempre y cuando la frontera de K sea suficientemente regular. En particular, si K es un cuerpo convexo cuya frontera es una variedad $(d-1)$-dimensional con curvatura de Gauss positiva (lo que llamaremos un *cuerpo convexo suave*) se tienen estimaciones asintóticas para $\hat{\chi}_R$ análogas a (**II**.7) y el problema de acotar α_K también se reduce a la estimación de una suma exponencial. Para esto último es común emplear el método de van der Corput [**38**], aunque para ciertas familias de cuerpos muy particulares la suma exponencial se conoce bien y cabe aplicar otras técnicas. Para cuerpos convexos suaves la conjetura (que debe tomarse con un poco de precaución) más extendida es $\alpha_K \leq 1/2$ para $d = 2$ (análogamente a lo que pasa con el círculo) y $\alpha_K = d-2$ para $d \geq 3$. Nuestro desconocimiento a la hora de tratar sumas exponenciales, sin embargo, hace que

para muy pocas familias de cuerpos se sepan obtener estos resultados. Por ejemplo, para las bolas y para los elipsoides racionales se ha probado la conjetura para $d \geq 4$, y para los irracionales si $d \geq 5$ [**37**]. El mejor resultado para $d = 2$ (el círculo) es el ya mencionado de Bourgain y Watt, y para $d = 3$ (la esfera) se sabe $\alpha_K \leq 21/16$, probado por Heath-Brown [**47**]. Para cuerpos convexos suaves en general las mejores cotas superiores conocidas son $\alpha_K \leq 131/208$ por Huxley [**59**], y $\alpha_K \leq d - 2 + r(d)$ con $r(d) = 78/158$ para $d = 3$ y $r(d) = (d^2 + 3d + 8)/(d^3 + d^2 + 5d + 4)$ para $d \geq 4$, ambos resultados por Guo [**39**]. Todo esto está contado con mucho más detalle en el capítulo 4 de esta memoria.

En el artículo [**15**] F. Chamizo muestra que para cuerpos convexos suaves tridimensionales que sean invariantes por rotaciones respecto al eje z basta con las estimaciones más sencillas de van der Corput para obtener $\alpha_K \leq 11/8$. Compárese con lo que se sabe para la esfera ($\alpha_K \leq 21/16$) y en general ($\alpha_K \leq 213/158$). Sin embargo, a la hora de obtener este resultado fue necesario imponer que la tercera derivada de la generatriz de K no se anulara en ningún punto. Más concretamente, si K es el sólido de revolución generado por rotación alrededor del eje z de la curva

$$\gamma(t) = \begin{cases} (t, 0, f_1(t)) & 0 \leq t \leq r_\infty \\ (2r_\infty - t, 0, f_2(2r_\infty - t)) & r_\infty \leq t \leq 2r_\infty \end{cases}$$



entonces se pedía que ninguna de las funciones $\frac{1}{r} f_i'''(r)$ (extendidas por continuidad a $r = 0$) se anulara en $0 \leq r < r_\infty$. Este tipo de condiciones aparecen a menudo al aplicar el método de van der Corput, y no suelen ser síntoma de ningún fenómeno subyacente inherente al problema en cuestión, sino simplemente resultado de nuestra incapacidad para entender bien dichas sumas. Con esta idea en mente nos propusimos F. Chamizo y yo eliminar, o al menos debilitar, la condición sobre $\frac{1}{r} f_i'''(r)$. Para ello el primer paso fue estudiar el caso más patológico: cuando $f_i'''(r)$ es idénticamente nula para $i = 1, 2$. La forma resultante es la del doble paraboloide de revolución

$$\{|z| \leq c - (x^2 + y^2)\}, \tag{Ⅱ.8}$$

para $c > 0$. Este cuerpo de revolución no tiene frontera suave: la frontera es singular en $z = 0$, pero aún así cumple la interesante propiedad de que la suma exponencial obtenida tras realizar sumación de Poisson es una versión truncada de la forma modular θ^2, donde θ es la función theta de Jacobi (Ⅱ.1). Esto permite usar una versión simplificada del método del círculo para dar cotas lo suficientemente fuertes sobre la suma exponencial como para deducir $\alpha_K \leq 1$:

Teorema. *Sea K el paraboloide de base elíptica $\{|z| \leq c - Q(\vec{x})\}$, donde Q es una forma cuadrática $(d-1)$-dimensional, definida positiva, cuya matriz $A = (a_{ij})$ cumple $a_{12}/a_{11}, a_{22}/a_{11} \in \mathbb{Q}$. Entonces $\alpha_K \leq d - 2$.*

La prueba de este resultado está contenida en el capítulo 5 de esta memoria, y en el artículo "Lattice points in elliptic paraboloids" [**20**] (conjunto con F. Chamizo).

Al comparar nuestro resultado con la literatura existente nos dimos cuenta de que el caso bidimensional (el de la doble parábola $\{|y| \leq c - x^2\}$) había sido resuelto 1975 por Popov [**81**], y su análogo en dimensión superior (**II**.8) había sido considerado por Krätzel en 1991 y 1997 [**71, 72**] obteniendo resultados más débiles que el del teorema enunciado. Hasta donde nos ha sido posible indagar, nuestro resultado proporciona el primer ejemplo de cuerpo tridimensional curvado para el cual se ha conseguido demostrar la conjetura. Tanto en el artículo de investigación como en el capítulo correspondiente de esta memoria aprovechamos para dar también algunos Ω-resultados más fuertes que los hasta ahora conocidos para casos particulares de la parábola y de los paraboloides en $d \geq 3$.

De vuelta al problema original concerniendo cuerpos de revolución convexos suaves, resultó que las técnicas utilizadas para el paraboloide eran demasiado particulares para ser inmediatamente aplicables al problema de debilitar la condición de no anulabilidad de $\frac{1}{r} f_i'''(r)$. Sin embargo si dan cierta intuición de qué ocurre cuando estas funciones tienen ceros de orden muy grande. Supongamos que los ceros de f_i''' son aislados. Si son de orden pequeño, entonces refinando los argumentos del artículo original de Chamizo [**15**] mediante una aplicación más enrevesada del método de van der Corput se recupera la cota $\alpha_K \leq 11/8$. Cuando los ceros son de orden mayor, la parte de la suma exponencial correspondiente a un entorno pequeño de la frontera de K cerca del cero de f_i''' resulta tener cierta aritmética (al fin y al cabo en esta zona la forma de K es muy similar a la de un paraboloide de revolución), y de nuevo podemos recuperar $\alpha_K \leq 11/8$ involucrando un argumento reminiscente al empleado para acotar la suma exponencial en el problema del paraboloide. Al final, mezclando ambos enfoques, obtenemos:

Teorema. *Supongamos que K es un cuerpo convexo, de frontera suave, curvatura de Gauss positiva e invariante por rotaciones alrededor del eje z. Si además las funciones generatriz f_i definidas arriba cumplen que los ceros de sus terceras derivadas f_i''' son de orden finito, entonces $\alpha_K \leq 11/8$.*

En particular este teorema abarca el caso de cuerpos con frontera analítica. Para probar este resultado dedicamos el capítulo 6 de esta memoria, contenido también incluido en el artículo "Lattice points in revolution bodies (II)" [**21**] (conjunto con F. Chamizo).

Acknowledgements

First and foremost, it is my advisor, Fernando Chamizo, who deserves the most credit. He is who introduced me to the wonderful topic of analytic number theory, and who spent countless afternoons explaining different aspects of this discipline and carefully reviewing each of my manuscripts. In fact, most of the ideas lying underneath this memoir were selflessly shared by him in one form or another. But above all this, he has also been a friend throughout the whole process of research and its inherent ups and downs.

During the few years that I stayed at the ICMAT (with a short visit to the MSRI) I also had the luck to stumble upon many interesting people. In particular, Ángel D. Martínez, Álvaro del Pino, Paco Torres and Corentin Perret-Gentil who were always eager to discuss any topic, be it about math, physics, computer science, biology or anything else. Without doubt they are first class mathematicians, scientists in general, and amazing friends.

Many others have also suffered from my constant visits to their offices in my spare (and not so spare) time. They are too many to be listed here, so let me extend my gratitude to you all. The atmosphere at the ICMAT could not have been better, and that is mostly due to the many personal interactions among all the predoc and postdoc students, with our different backgrounds and interests. I truly hope it will stay this way for years to come.

Some particular people deserve a special mention. I have shared many good memories in the company of Ángel, Eric Latorre, Mª Ángeles García, Tania Pernas and Diego Alonso. In particular, the latter two hosted the merriest Christmas dinners one could ever imagine. I also have fond memories of the incredible gastronomic tour I embarked on with Carlos Vinuesa and Joan Tent, and the short-lived but intense "beer fridays" that I can blame on Manuel Jesús Pérez and Antonio Pérez.

Many people have also made my travels to other places feel a little warmer. Michael Elie (and his lovely kids) in Berkeley, Èlia Casas and Alex Sanglas in Barcelona, Pablo Portilla in Bilbao, Rocío Saavedra in Murcia, Olgierd Borowiecki in Göttingen and Alba Delgado, without whom my visit to Sevilla would have not been as magic.

During my PhD I also had the chance to collaborate with Iason Efraimidis on an article which was finally not included in this dissertation. It has been a very gratifying experience and the article really benefited from his elegance.

Hay algunos amigos a los que me es imposible reservarles el espacio que se merecen. Entre ellos, Iris Valero, que un haiku del destino ha traído de vuelta a mi vida. Tampoco puedo dejar de lado a Aury Belmonte, Agustín Ruiz-Escribano, Ismael Díaz, Marién López ni Daniel Martínez, que llevan tiempo acompañándome en este viaje que es la vida.

Quiero dedicar esta memoria, así como todo el trabajo que hay detrás, muy especialmente a mi familia. Porque les debo a ellos el haber podido siempre centrarme en lo que realmente me gusta y me motiva. Ellos me han allanado el camino, y me han alentado a no abandonar nunca. A mis padres, ante todo, por estimularme desde pequeño, y por estar siempre ahí.

It goes without saying that both me and this dissertation greatly benefited from the influence of a handful of excelent professors that I had during my education. I must also thank Daniel Bump for carefully reading my first article and providing many suggestions. Corentin for the fruitful conversations at the MSRI leading to the result on lattice points in revolution bodies. Ángel reviewed several early manuscripts and provided many useful suggestions. Chantal David made possible my stay at Berkeley by selflessly providing me with the necessary paperwork, even though we had not met before, and the staff at the MSRI who were absolutely helpful during my visit.

I would also like to take the opportunity to thank the thesis committee and the researchers that were asked to write reports, for all the effort involved in reading this material and the useful suggestions received.

This work would have not been possible without the financial aid provided by "la Caixa" through their "la Caixa"-Severo Ochoa international PhD programme at the Instituto de Ciencias Matemáticas (CSIC-UAM-UC3M-UCM), and later on by the unemployment benefits program of the government of Spain.

Last but not least, thanks to you, the reader, for without you the effort put into writing this dissertation would make no sense.

List of symbols

Vinogradov-Landau-Hardy notation:

$f \ll g$	$\lvert f(x)\rvert \leq C\lvert g(x)\rvert$ for some nonzero constant C, specially in the neighborhood of a point.
$f \gg g$	Same as $g \ll f$.
$f \asymp g$	We have $f \ll g \ll f$, *i.e.* $C_1\lvert g(x)\rvert \leq \lvert f(x)\rvert \leq C_2\lvert g(x)\rvert$ for nonzero constants C_1 and C_2.
$f \sim g$	Neither f nor g vanish in the neighborhood of a point and $\lim f/g = 1$.
$f = O(g)$	Same as $f \ll g$.
$f = o(g)$	If g does not vanish, $\lim f/g = 0$. In general, this means we can write $f = gh$ for some function h with $\lim h = 0$.
$f = \Omega(g)$	The negation of $f = o(g)$. In other words, for some constant $C > 0$ one has $\lvert f(x)\rvert \geq C\lvert g(x)\rvert$ for infinitely many values of x close to a certain point.[1]
$f = \Omega_+(g)$	Equivalent to $\max\big(f(x), 0\big) = \Omega(g)$.
$f = \Omega_-(g)$	Equivalent to $\min\big(f(x), 0\big) = \Omega(g)$.
ϵ	An arbitrarily small quantity which may vary from instance to instance.

Functions related to the fractional part of a number:

$\lfloor x \rfloor$	Integer part of x, *i.e.* biggest integer n satisfying $n \leq x$.
$\lceil x \rceil$	Ceil of x, *i.e.* smallest integer n satisfying $n \geq x$.
$\{x\}$	Decimal part of x, equivalent to $x - \lfloor x \rfloor$.
$\lVert x\rVert_{\mathbb{Z}}$	Distance from x to the nearest integer.
$\psi(x)$	Saw-tooth function $\psi(x) = x - \lfloor x \rfloor - 1/2$.[2]
$e(x)$	Equivalent to $\exp\big(2\pi i x\big)$.

Other symbols:

$\lVert\cdot\rVert_p$	p-norm of either a vector or a function.
$\lVert\cdot\rVert$	2-norm of either a vector or a function. Equivalent to $\lVert\cdot\rVert_2$.
$\vec{v}^t$ or A^t	Transpose of either the vector $\vec{v}$ or the matrix A.
$\vec{v}\cdot\vec{w}$	Inner product between $\vec{v}$ and $\vec{w}$.
$\#\Omega$	Cardinality of the set Ω.
$:=$	Left hand side is defined as the right hand side.

[1] Not to be confused with Knuth's version widely used in computer science.

[2] The symbol ψ is also used to denote a wavelet in §3.4.

$\theta(x)$	Jacobi's theta function, defined by (**I**.1).
$\varphi(x)$	"Riemman's nondifferentiable example", defined by (**I**.9).
$\mathbb{H}$	The upper half-plane $\{z \in \mathbb{C} \,:\, \Im z > 0\}$.
$\mathbb{F}$ or $\mathbb{F}_\Gamma$	Fundamental domain of either $\mathrm{SL}_2(\mathbb{Z})$ or Γ. See §1.2, §2.3.
$\mathcal{F}_x(\delta)$	Speiser circle over x of radius δ, defined in §1.4.
$\mathcal{A}_x$	Interval associated to x in a Farey dissection, see §1.5.
$\mathcal{N}(R)$	Number of lattice points in RK, see §4.1.
α_K	Error exponent $\inf\{\alpha : \mathcal{N}(R) - \lvert RK\rvert = O(R^\alpha)\}$, see §4.1.
$\mathrm{GL}_n(\mathcal{R})$	Group of invertible $n \times n$ matrices over the ring $\mathcal{R}$.
$\mathrm{GL}_n^+(\mathbb{R})$	Group of invertible $n \times n$ matrices over the real numbers with positive determinant.
$\mathrm{SL}_n(\mathcal{R})$	Group of $n \times n$ matrices with determinant equal to 1 over the ring $\mathcal{R}$.

All the remaining symbols are either standard or locally defined in the same section or chapter where they are used.

Bibliography

[1] T. Asai. *On the Fourier coefficients of automorphic forms at various cusps and some applications to Rankin's convolution.* J. Math. Soc. Japan, 28(1):48–60, 1976.

[2] A. O. L. Atkin, J. Lehner. *Hecke operators on* $\Gamma_0(m)$. Math. Ann., 185:134–160, 1970.

[3] F. Bars. *The group structure of the normalizer of* $\Gamma_0(N)$. arXiv:math/0701636v1.

[4] P. T. Bateman, S. Chowla, and P. Erdös. *Remarks on the size of* $L(1,\chi)$. Publ. Math. Debrecen, 1:165–182, 1950.

[5] V. Bentkus, F. Götze. *On the lattice point problem for ellipsoids.* Acta Arith., 80(2):101–125, 1997.

[6] V. Bentkus, F. Götze. *Lattice point problems and distribution of values of quadratic forms.* Ann. of Math. (2), 150(3):977–1027, 1999.

[7] V. Beresnevich, F. Ramirez, S. Velani. *Metric Diophantine Approximation: aspects of recent work.* In Dynamics and Analytic Number Theory, chapter 1, Cambridge Univ. Press, 2016.

[8] V. Blomer. *Uniform bounds for Fourier coefficients of theta-series with arithmetic applications.* Acta Arith., 114(1):1–21, 2004.

[9] P. Du Bois-Reymond. *Versuch einer Classification der willkürlichen Functionen reeller Argumente nach ihren Aenderungen in den kleinsten Intervallen.* J. Reine Angew. Math. 79:21–37, 1875.

[10] E. Bombieri, H. Iwaniec. *On the order of* $\zeta(\frac{1}{2}+it)$. Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4), 13(3):449–472, 1986.

[11] J. Bourgain and E. Watt. *Mean square of zeta function, circle problem and divisor problem revisited.* arXiv:1709.04340, 2017.

[12] H. Bremermann. *Distributions, Complex Variables and Fourier Transforms.* Addison-Wesley Series in Mathematics, Addison-Wesley Publishing Company, 1965.

[13] P. I. Butzer, E. I. Stark. *"Riemann's example" of a continous nondifferentiable function in the light of two letters (1865) of Christoffel to Prym.* Bull. Soc. Math. Belg., 38:45–73, 1986.

[14] F. Chamizo. *Automorphic Forms and Differentiability Properties.* Trans. Amer. Math. Soc., 356(5):1909–1935 (electronic), 2004.

[15] F. Chamizo. *Lattice points in bodies of revolution.* Acta Arith., 85(3):265–277, 1998.

[16] F. Chamizo, E. Cristóbal, A. Ubis. *Lattice points in rational ellipsoids.* J. Math. Anal. Appl., 350(1):283–289, 2009.

[17] F. Chamizo, H. Iwaniec. *On the Sphere Problem.* Rev. Mat. Iberoamer., 11:417–429, 1995.

[18] F. Chamizo, H. Iwaniec. *On the Gauss mean-value formula for class number.* Nagoya Math. J., 151:199–208, 1998.

[19] F. Chamizo, I. Petrykiewicz, S. Ruiz-Cabello. *The Hölder exponent of some Fourier series.* J. Fourier Anal. Appl., 23(4):758–777, 2017.

[20] F. Chamizo, C. Pastor. *Lattice points in elliptic paraboloids.* arXiv:1611.04498v2, 2017 (to appear in Publicacions Matemàtiques).

[21] F. Chamizo, C. Pastor. *Lattice points in bodies of revolution II.* arXiv:1709.08593v2, 2017.

[22] H. Cohn. *Advanced Number Theory.* Dover Publications Inc., 1980.

[23] H. Davenport. *Multiplicative Number Theory.* Springer-Verlag, 2000.

[24] J. J. Duistermaat. *Selfsimilarity of "Riemann's Nondifferentiable Function".* Nieuw Arch. Wisk., 9(3):303–337, 1991.

[25] W. Duke. *An introduction to the Linnik problems.* Chapter in *Equidistribution in Number Theory, An Introduction*, Springer, 2007.

[26] W. Duke. *Lattice points on ellipsoids.* Sém. Théor. Nombres Bordeaux (2), 1987–88. No. 37, pp. 1–6.

[27] W. Duke, R. Schulze-Pillot. *Representation of integers by positive ternary quadratic forms and equidistribution of lattice points on ellipsoids.* Invent. Math., 99(1):49–57, 1990.

[28] M. Eichler. *Eine Verallgemeinerung der Abelschen Integrale.* Math. Z. 67:267–298, 1957.

[29] H. Fiedler, W. Jurkat, and O. Körner. *Asymptotic expansions of finite theta series.* Acta Arith., 32(2):129–146, 1977.
[30] J. R. Ford. *Fractions.* Amer. Math. Monthly, 45(9):586–601, 1938.
[31] R. Fricke, F. Klein. *Vorlesungen über die Theorie der elliptischen Modulfunktionen.* 2 Bände, Teubner-Verlag, 1890.
[32] F. Fricker. *Einführung in die Gitterpunktlehre.* Volume 73 of *Mathematische Reihe*, Birkhäuser Verlag, 1982.
[33] U. Frisch, G. Parisi. *On the singularity structure of fully developed turbulence.* Appendix in *Fully Developed Turbulence and Intermittency*, in *Turbulence and predictability in geophysical fluid dynamics and climate dynamics*, Varenna, 1983, Proceedings of the International School of Physic Enrico Fermi, North-Holland, 1985.
[34] K. F. Gauss. *De nexu inter multitudinem classium, in quas formae binariae secundi gradus distribuntur, earumque determinantem.* In *Werke*, volume 2, 269–291. Georg Olms Verlag, Hildesheim, 1981.
[35] J. Gerver. *The differentiability of the Riemann function at certain rational multiples of π.* Amer. J. Math., 92:33–55, 1970.
[36] J. Gerver. *More on the differentiability of the Riemann function.* Am. J. Math., 93(1):33–41, 1971.
[37] F. Götze. *Lattice point problems and values of quadratic forms.* Invent. Math., 157(1):195–226, 2004.
[38] S. W. Graham, G. Kolesnik. *Van der Corput's method of Exponential Sums.* Volume 126 of *London Mathematical Society Lecture Note Series*, Cambridge University Press, Cambridge, 1991.
[39] J. Guo. *On lattice points in large convex bodies.* Acta Arith., 151(1):83–108, 2012.
[40] J. L. Hafner. *New omega theorems for two classical lattice point problems.* Invent. Math., 63(2):181–186, 1981.
[41] G. H. Hardy. *On the Expression of a Number as the Sum of Two Squares.* Quart. J. Math., 46:263–283, 1915.
[42] G. H. Hardy. *Weierstrass's nondifferentiable function.* Trans. Amer. Math. Soc., 17(3):301–325, 1916.
[43] G. H. Hardy. *The average order of the functions $P(x)$ and $\Delta(x)$.* Proc. London Math. Soc., 15(2):192–213, 1916.
[44] G. H. Hardy, J. E. Littlewood. *Some problems of diophantine approximation I: The fractional part of $n^k\theta$.* Acta Math., 37:155–191, 1914.
[45] G. H. Hardy, J. E. Littlewood. *Some problems of diophantine approximation II: The trigonometrical series associated with the elliptic ϑ-functions.* Acta Math., 37:193–239, 1914.
[46] G. H. Hardy, E. M. Wright. *An introduction to the Theory of Numbers.* Oxford University Press, 2008.
[47] D. R. Heath-Brown. *Lattice points in the sphere.* In *Number theory in progress*, 883–892, de Gruyter, Berlin, 1999.
[48] D. R. Heath-Brown. *Ternary quadratic forms and sums of three square-full numbers.* In *Séminaire de Théorie des Nombres, Paris 1986–87*, volume 75 of *Progr. in Math.*, Birkhäuser Boston, 1988.
[49] K. Henriot, K. Hughes. *On restriction estimates for discrete quadratic surfaces.* arXiv:1611.00720, 2016.
[50] C. S. Herz. *Fourier Transform Related to Convex Sets.* Annals of Mathematics, Second Series, 75(1):81–92, 1962.
[51] E. Hlawka. *Über Integrale auf konvexen Körpern I.* Monatsh. Math. 54:1–36, 1950.
[52] M. Holschneider, Ph. Tchamitchian. *Pointwise analysis of Riemann's "nondifferentiable" function.* Invent. Math., 105(1):157–175, 1991.
[53] L. Hörmander. *The Analysis of Partial Differential Operators I.* Springer-Verlag Berlin Heidelberg, 2003
[54] F. de la Hoz, L. Vega. *Vortex filament equation for a regular polygon.* Nonlinearity, 27(12):3031–3057, 2014.
[55] L.-K. Hua. *The lattice-points in a circle.* Quart. J. Math., Oxford Ser., 13:18–29, 1942.
[56] L.-K. Hua. *Introduction to number theory.* Springer-Verlag, Berlin, 1982.
[57] M. N. Huxley. *Area, lattice points and exponential sums.* In *Proceedings of the International Congress of Mathematicians, Vol. I, II (Kyoto, 1990)*, pp. 413–417, Math. Soc. Japan, 1991.

[58] M. N. Huxley. *Area, lattice points, and exponential sums*, volume 13 of *London Mathematical Society Monographs. New Series*, Oxford University Press, 1996.

[59] M. N. Huxley. *Exponential sums and lattice points. III.* Proc. London Math. Soc. (3), 87(3):591–609, 2003.

[60] A. Ivić, E. Krätzel, M. Kühleitner, W. G. Nowak. *Lattice points in large regions and related arithmetic functions: recent developments in a very classic topic.* In *Elementare und analytische Zahlentheorie*, 89–128, Franz Steiner Verlag Stuttgart, 2006.

[61] H. Iwaniec. *Topics in Classical Automorphic Forms.* Vol. 17 of *Graduate Studies in Mathematics*, Amer. Math. Soc., 1997.

[62] H. Iwaniec, E. Kowalski. *Analytic number theory.* Colloquium publications (Amer. Math. Soc.), 2004.

[63] C. G. J. Jacobi. *Fundamenta nova theoriae functionum ellipticarum.* Königsberg, 1829. Reprinted by Cambridge University Press, 2012.

[64] S. Jaffard. *The spectrum of singularities of Riemann's function.* Rev. Mat. Iberoamericana, 12(2):441–460, 1996.

[65] S. Jaffard. *Local behavior of Riemann's function.* In *Harmonic analysis and operator theory (Caracas, 1994)*, volume 189 of *Contemp. Math.*, pp. 287–307. Amer. Math. Soc, 1995.

[66] V. Jarník. *Über die simultanen diophantischen Approximationen.* Math. Z., 33(1):505–543, 1931.

[67] A. Kar. *Weyl's Equidistribution Theorem.* Resonance, 8(5):30–37, 2003.

[68] A. Ya. Khinchin. *Continued fractions.* Dover Publications Inc., 1997.

[69] N. Koblitz. *Introduction to elliptic curves and modular forms.* Springer-Verlag, 1984.

[70] G. Kolesnik. *On the order of $\zeta(\frac{1}{2}+it)$ and $\Delta(R)$.* Pacific J. Math., 98(1):107–122, 1982.

[71] E. Krätzel. *Lattice points in elliptic paraboloids.* J. Reine Angew. Math., 416:25–48, 1991.

[72] E. Krätzel. *Weighted lattice points in three-dimensional convex bodies and the number of lattice points in parts of elliptic paraboloids.* J. Reine Angew. Math., 485:11–23, 1997.

[73] E. Landau. *Neue Untersuchungen über die Pfeiffersche Methode zur Abschätzung von Gitterpunktanzahlen.* Sitzungsber. d. math-naturw. Classe der Kaiserl. Akad. d. Wissenschaften, 2. Abteilung, Wien, 124:469–505, 1915.

[74] D. H. Lehmer. *Incomplete Gauss sums.* Mathematika, 23(2):125–135, 1976.

[75] J. E. Littlewood. *On the Class-Number of the Corpus $P(\sqrt{-k})$.* Proc. London Math. Soc., S2-27(1):358, 1928.

[76] The LMFDB Collaboration. *The L-functions and Modular Forms Database.* `http://www.lmfdb.org`, 2013. [Online; accessed 4 March 2016].

[77] S. D. Miller, W. Schmid. *The Highly Oscillatory Behavior of Automorphic Distributions for $\mathrm{SL}_2(\mathbb{Z})$.* Letters in Math. Physics, 69(1):265–286, 2004.

[78] H. L. Montgomery. *Ten lectures on the interface between analytic number theory and harmonic analysis*, volume 84 of *CBMS Regional Conference Series in Mathematics*. Published for the Conference Board of the Mathematical Sciences, by the American Mathematical Society, 1994.

[79] L. Mordell. *On the Kusmin-Landau inequality for exponential sums.* Acta Arith., 4(1):3–9, 1958.

[80] C. Pastor. *On the regularity of fractional integrals of modular forms.* arXiv:1603.06491, 2016 (to appear in Trans. of the Amer. Math. Soc.).

[81] V. N. Popov. *The number of lattice points under a parabola.* Mat. Zametki, 18(5):699–704, 1975.

[82] R. A. Rankin. *Modular forms and functions.* Cambridge Univ. Press, 1977.

[83] S. Ruiz-Cabello. *Generadores de primos, identidades aproximadas y funciones multifractales.* PhD dissertation, Universidad Autónoma de Madrid, 2014.

[84] *SageMath, the Sage Mathematics Software System (Version 8.0)*, The Sage Developers, 2017, `http://www.sagemath.org`.

[85] J-P. Serre. *A Course in Arithmetic.* Springer-Verlag, 1973.

[86] S. Seuret, J. L. Véhel. *The local Hölder function of a continuous function.* Appl. Comput. Harmon. Anal., 13(3):263–276, 2002.

[87] C. L. Siegel. *Lectures on quadratic forms.* Notes by K. G. Ramanathan. Lectures on Mathematics, no. 7, Tata Institute of Fundamental Research, Bombay, 1967.

[88] C. L. Siegel. *The average measure of quadratic forms with given determinant and signature.* Ann. of Math., 45(2):667–685, 1944.

[89] W. Sierpiński. *Sur la sommation de la série $\sum_{n>a}^{n\leq b}\tau(n)f(n)$, où $\tau(n)$ signifie le nombre des décompositions du nombre n en une somme de deux carrés de nombres entiers.* In *Oeuvres Choisies*, PWN - Êditions Scientifiques de Pologne, 1974.

[90] K. Soundararajan. *Omega results for the divisor and circle problems.* Int. Math. Res. Not., 36:1987–1998, 2003.

[91] S. L. Velani. *Diophantine approximation and Hausdorff dimension in Fuchsian groups.* Math. Proc. Cam. Phil. Soc., 113:343–354, 1993.

[92] G. Voronoï. *Sur une fonction transcendante et ses applications à la sommation de quelques séries.* Ann. scient. de l'École Normale supè., 21:203–267 and 459–533, 1904.

[93] G. Voronoï. *Sur le développment à l'aide des fonctions cylindriques, des sommes doubles* $\sum f(pm^2 + 2qmn + 2n^2)$*, où* $pm^2 + 2qmn + 2n^2$ *est une forme positive à coefficients entiers.* Proceedings of the *Verh. III Intern. Math. Kongr. Heidelberg (1904)*, pp. 241–245, Leipzig, 1905.

[94] G. N. Watson. *A treatise on the Theory of Bessel Functions.* Cambridge University Press, 1996.

[95] K. Weierstrass. *Über continuierliche Functionen eines reellen Arguments, die für keinen Werth des letzteren einen bestimmten differentialquotienten besitzen.* In *Mathematische Werke II*, pp. 71-74. Königl. Akad. Wiss., 1872.

[96] E. T. Whittaker, G. N. Watson. *A course in modern analysis.* Cambridge Univ. Press, 1915.

[97] D. Zagier. *Elliptic modular forms and their applications.* Chapter in *The 1-2-3 of Modular Forms*, Springer-Verlag Berlin Heidelberg, 2008.

[98] A. Zygmund. *Trigonometric series.* Vol I, II. Cambridge Univ. Press, 2002.